REPORT RESUMES

ED 003 482 24
INDIVIDUAL DIFFERENCES IN LEARNING--INTERFERENCE FACTOR.
BY- JENSEN, ARTHUR R.
CALIFORNIA UNIV., BERKELEY, INST.OF HUMAN LEARNING
REPORT NUMBER CRP-1867 PUB DATE 64
REPORT NUMBER BR-5-0822
EDRS PRICE MF-$0.27 HC-$6.76 169P.

DESCRIPTORS- *PSYCHOMETRICS, *INDIVIDUAL DIFFERENCES, *LEARNING, *COLLEGE STUDENTS, *INTELLIGENCE, FACTOR ANALYSIS, BERKELEY, CALIFORNIA

AN INVESTIGATION WAS CONDUCTED (1) TO DEVISE A NUMBER OF LABORATORY MEASURES OF INDIVIDUAL DIFFERENCES IN SUSCEPTIBILITY TO INTERFERENCE EFFECTS IN LEARNING AND (2) TO DETERMINE THEIR DIMENSIONALITY IN A VARIETY OF INTERFERENCE EFFECTS. SUBJECTS WERE 530 STUDENTS IN INTRODUCTORY COURSES IN EDUCATIONAL PSYCHOLOGY AT THE UNIVERSITY OF CALIFORNIA, BERKELEY. TWO TYPES OF EXPERIMENTAL LEARNING TASKS WERE USED--(1) SHORT-TERM MEMORY (STM) AND (2) SERIAL LEARNING. THE STM INCLUDED (1) IMMEDIATE DIGIT SPAN, (3) IMMEDIATE AND DELAYED COLOR-FORM MEMORY SPAN, (5) PROACTIVE INHIBITION OF DIGIT SPAN, (6) FORWARD AND BACKWARD DIGIT SPAN, PRE- AND POST-CUED, AND (7) REPEATED DIGIT SPAN. THE SERIAL TASKS INVOLVED (1) RETROACTIVE INHIBITION OF SERIAL LEARNING AND (2) DIFFERENT AMOUNTS OF INTRALIST INTERFERENCE. AGE, INTELLIGENCE (AS MEASURED BY "RAVEN'S PROGRESSIVE MATRICES"), COLLEGE GRADE-POINT AVERAGE, AND TWO PERSONALITY MEASURES WERE INCLUDED AS VARIABLES. COLLEGE GRADE-POINT AVERAGE WAS MOST HIGHLY CORRELATED WITH THE FACTOR OF RESISTANCE TO RESPONSE COMPETITION. (JK)

5-0822

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Individual Differences in Learning: Interference Factor

Cooperative Research Project No. 1867

Arthur R. Jensen

Institute of Human Learning
University of California
Berkeley, California

1964

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

ED 003482

Table of Contents

Page

I. Statement of the Problem . . . 1
  A. General Background of the Present Study . . . 2
    1. The Basic Problem . . . 2
    2. Phenotypes and Genotypes . . . 4
  B. Approaches to the Study of IDs in Learning . . . 6
    1. Intrinsic and Extrinsic IDs . . . 6
    2. The Psychometric Approach . . . 7
  C. Learning Ability and Intelligence . . . 10
  D. The Definition of Learning . . . 11
  E. Innate and Environmental Sources of IDs in Learning . . . 12
  F. The Approach of the Present Study . . . 13

II. Objectives and Design of the Experiment . . . 13
  A. Underlying Philosophy of This Research . . . 13
  B. Test Selection and Development . . . 17
    1. Rationale of Test Selection . . . 19
    2. Establishing the Hypothesized Effect of the Experimental Variable . . . 23
    3. Establishing Valid Individual Differences in the Hypothesized Effect . . . 23
    4. Establishing Reliability of Measurement. . . . 24
  C. Factor Analysis of the Tests . . . 25

III. Related Research . . . 25
  A. Distinction Between Long-Term and Short-Term Memory . . . 25
  B. Interference in Immediate Memory Span . . . 29
  C. Interference and Individual Differences . . . 31
    1. Age . . . 31
    2. Interference in Slow and Fast Learners . . . 33

IV. Procedures . . . 34
  A. Subjects . . . 34
  B. Apparatus . . . 36
  C. Unsatisfactory and Abandoned Techniques . . . 37
    1. Retroactive Inhibition of Selective Trial-and-S-R Learning . . . 38
    2. Intra-task Interference in Selective Trial-and-Error S-R Learning . . . 40
    3. Interference Due to S-R Similarity . . . 42
  D. Interference in Short-term Memory Tasks . . . 43
    1. Immediate Digit Span and Delayed Digit Span . . . 44
    1A. Auditory Form of the Immediate and Delayed Memory Span Tests . . . 46
    2. Immediate and Delayed Color-Form Span . . . 46
    3. Retroactive and Proactive Inhibition of Digit Span . . . 47
    4. Forward and Backward Digit Span . . . 49
    5. Digit Span and Repeated Digit Span . . . 50
  E. Interference in Serial Learning . . . 51
    1. Serial Intralist Interference . . . 52
    2. Serial Retroactive Interference . . . 53
  F. Interference From Response Competition: The Stroop Test . . . 54

Page

G. Personality Inventories 55
H. Ancillary Data 56
I. Order of Administering Tests 56

V. Analysis and Interpretation of the Data 57
A. Memory Span Tests 57
Scoring Methods 57
1. Immediate and Delayed Digit Span 60
a. Serial Position Effects in Recall 60
b. Effects of Interference and Delay in Recall . . . 63
c. Effects of Practice 63
d. Correlation Between Scoring Methods 64
e. Descriptive Statistics of Ss' Performance 65
f. Analysis of Variance 66
g. Reliability of Scores 66
h. Correlational Analysis of Scores 67
1A. Auditory Form of the Immediate and Delayed Memory Span Tests 68
a. Serial Position Effects in Recall 68
b. Effects of Interference and Delay in Recall . . . 68
c. Effects of Practice 70
d. Correlation Between Scoring Methods 70
e. Descriptive Statistics of Ss' Performance 70
f. Analysis of Variance 71
g. Reliability of Scores 71
h. Correlational Analysis of Scores 71
i. Correlation Between Auditory and Visual Performance 71
2. Immediate and Delayed Color-Form Span 72
a. Serial Position Effects in Recall 72
b. Effects of Interference and Delay in Recall . . . 75
c. Effects of Practice 75
d. Correlation Between Scoring Methods 76
e. Descriptive Statistics of Ss' Performance 76
f. Analysis of Variance 76
g. Reliability of Scores 76
h. Correlational Analysis of Scores 77
3. Retroactive and Proactive Inhibition of Digit Span . 77
a. Serial Position Effects in Recall 78
b. Effects of RI, PI, and Delay in Recall 79
c. Effects of Practice 80
d. Correlation Between Scoring Methods 80
e. Descriptive Statistics of Ss' Performance 80
f. Analysis of Variance 80
g. Reliability of Scores 81
h. Correlational Analysis of Scores 82
4. Forward and Backward Digit Span 85
a. Serial Position Effects in Recall 85
c. Effects of Practice 86
d. Correlation Between Scoring Methods 86
e. Descriptive Statistics of Ss' Performance 87
f. Analysis of Variance 87
g. Reliability of Scores 88
h. Correlational Analysis of Scores 88

Page

- 5. Digit Span and Repeated Digit Span 90
    - a. Serial Position Effects in Recall 90
    - b. Effect of Repetition in Recall 92
    - c. Effects of Practice 92
    - d. Correlation Between Scoring Methods 92
    - e. Descriptive Statistics of Ss' Performance 93
    - f. Analysis of Variance 93
    - g. Reliability 93
    - h. Correlational Analysis of Scores 94
- B. Serial Learning Tasks 94
    - Scoring Methods 94
    - 1. Serial Intralist Interference 96
        - a. Serial Position 96
        - b. Effects of Practice 96
        - c. Descriptive Statistics of Ss' Performance . . . 98
        - d. Reliability 98
        - e. Correlational Analysis of Scores 99
    - 2. Serial Retroactive Interference 100
        - a. Serial Position Effects 100
        - b. Effects of Practice 101
        - c. Descriptive Statistics of Ss' Performance . . . 101
        - d. Reliability of Scores 102
        - e. Correlational Analysis of Scores 103
- C. The Stroop Color-Word Test 105
    - a. Effects of Practice 108
    - b. Descriptive Statistics of Ss' Performance 109
    - c. Reliability of the Scores 109
    - d. Correlations of Stroop With Other Variables . . . 109
- D. Factor Analysis of All Tests 111
    - 1. Principal Components Analysis 115
    - 2. Rotated Oblique First Order Factors 119
    - 3. Rotated Oblique Second Order Factors 131
    - 4. Rotated Oblique Third Order Factors 138
    - 5. General Factor 141

VI. Summary, Conclusions, and Implications 141

VII. References 149

VIII. Appendices 154
- A. Analysis of Variance of Visual and Auditory Memory Span Tests 154
- B. Analysis of Variance of Color-Form Memory Span Tests . . 155
- C. Analysis of Variance of Forward-Backward Memory Span Tests 156
- D. Analysis of Variance of Digit Span and Repeated Digit Span Tests 157
- E. Factor Loadings Corrected for Attenuation 158
- F. Estimated Standard Error of Factor Loadings After Correction for Attenuation 159
- G. Correlations of the Maudsley Personality Inventory With Memory Span Tests in the Single-Test Groups 160

List of Tables

Page

Table 1A. Probability of Recall in Visual Digit Series (IDS, 2-9 and DDS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each p Value Based on 1000 Observations (100 Ss X 10 Replications). . . . 61

Table 1B Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 1A), as a Function of Length of Series and Serial Position (Visual Presentation). 62

Table 2A Probability of Recall in Auditory Digit Series Tests (IDS, 2-9 and DDS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each p Value Based on 750 Observations (50 Ss X 10 Replications + 50 Ss X 5 Replications). 69

Table 2B Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 2A), as a Function of Length of Series and Serial Position (Auditory Presentation). 69

Table 3A Probability of Recall in Color-Form Series (ICFS, 2-9) and DCFS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each p Value Based on 550 Observations (50 Ss X 5 Replications + 30 Ss X 10 Replications). 73

Table 3B Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 3A), as a Function of Length of Series and Serial Position (Color-Form Series). 74

Table 4A Probability of Recall in Proactive and Retroactive Digit Series Tests (DS-RI,4-7; DS-PI,4-7; DDS,4-7; IDS,4-7) as a Function of Length of Series, Serial Position, and Experimental Condition (Retroactive Interference, Proactive Interference, Delayed Recall, Immediate Recall. Each p Value Based on 624 Observations (52 Ss X Replications + 52 Ss X 4 Replications). 78

Table 4B Amount of Retroactive Interference (RI) in Terms of Decrement in Probability of Recall (D-RI in Table 4A), Proactive Interference (D-PI in Table 4A), and Decrement Due to Delay in Recall (I-D in Table 4A), as a Function of Series Length and Serial Position. 79

Table 5 Analysis of Variance of Total Recall Scores in Conditions RI and PI. 81

Table 6 Intercorrelations (Pearson r) Among RI, PI, Immediate, and Delayed Recall Scores for Single-Tests Group (N = 52). (Day 1 correlations above the diagonal, Day 2 below.). 82

Tables Cont. Page

Table 7 Intercorrelations Among Variables in Table 6 for Days 1 and 2 Combined. 82

Table 8 Intercorrelations Among RI, PI, D, and I for All-Tests Group (N = 50). 82

Table 9 Probability of Recall in Digit Series Tests as a Function of Serial Position, Order of Recall (Forward vs. Backward), and Time S was Informed of Order of Recall (Pre-Cued vs. Post-Cued). Each p Value Based on 1760 Observations (30 Ss X 32 Replications + 50 Ss X 16 Replications). 85

Table 10 Means and SDs of Percentage of Maximum Possible Score in the Various Conditions of the Forward and Backward Digit Span Tests. (Each Mean and SD Based on 30 Ss X 16 Replications.). 87

Table 11 Reliability of Position and Sequence Scores on the Forward (F) and Backward (B) Digit Span Test Under Pre and Post Cuing for Direction of Recall. 88

Table 12A Probability of Recall in Digit Series Tests (DS, 5-12, and RDS, 5-12) as a Function of Length of Series, Serial Position, and Number of Presentations of Series (1 vs. 2). Each p Value Based on 550 Observations (30 Ss X 10 Replications + 50 Ss X 5 Replications). 91

Table 12B Amount of Increment in Probability of Recall Due to One Repetition of Series (i.e., 2-1 in Table 12A), as a Function of Length of Series and Serial Position. 91

Table 13 Means and SDs of Errors as a Function of Serial Position for Lists Composed of Words (W), Trigrams (Tri), and Color-Forms (CF), and the Mean Correlation Between Ss' Serial Position Curves. (Single-Test Group, N = 30). 97

Table 14 Mean Errors and Index of Oscillation for the Set of Serial Intralist Interference Tests. 98

Table 15 Means and SDs of Total Errors and the Index of Oscillation for the Various Lists of the Serial Retroaction Paradigm. . 102

Table 16 Test-Retest (Equivalent Forms) and Internal Consistency Reliabilities of the Various Lists of the Serial Retroaction Paradigm. 103

Table 17 Intercorrelations Among Variables in the IL-High and IL-Low Groups (N = 25 in each group) for the Serial Retroactive Inhibition Paradigm. (Day 1 above the diagonal, Day 2 below) 104

Table 18 Intercorrelations Among Variables in the All-Tests Group (N = 50) for the Serial Retroactive Inhibition Paradigm. . 104

Tables Cont. Page

Table 19 Basic Stroop Scores and Scoring Formulas. 106

Table 20 Principal Components and Varimax Rotation of Stroop Scores (First + Second Administration) ($\underline{N}$ = 436). 107

Table 21 Analysis of Variance of Basic Stroop Scores (A, B, C) for 10 Administrations. 108

Table 22 Correlations of Stroop Factors with Memory Span Tests (Position Scores) in the Single-Test and All-Tests Groups (Decimals omitted). 110

Table 23 Correlations of Stroop Factors with Serial Learning Tests in the All-Tests Group ($\underline{N}$ = 50) (Decimals omitted). 112

Table 24 Means, SDs, and Test-Retest Reliabilities of the Variables Included in the Factor Analysis of the All-Tests Group ($\underline{N}$ = 50). 113

Table 25 Correlations Among Variables in the All-Tests Group ($\underline{N}$ = 50) (Decimals omitted). 116-116a

Table 26 Principal Components and Communalities ($\underline{h}^2$) of the Variables in the All-Tests Group (Decimals omitted). 117

Table 27 Totated Oblique Factor Coefficients of First-Order Factors (Decimals omitted). 121

Table 28 Correlations of Variables With Oblique Factors (Decimals omitted). 122

Table 29 Intercorrelations Among the Rotated Oblique Factors[1] (Table 27) (Decimals omitted). 132

Table 30 Rotated Oblique Second-Order Factor Coefficients (Decimals omitted). 133

Table 31 Correlations of Variables With Oblique Second-Order Factors (Decimals omitted). 134

Table 32 Correlations Among the Second-Order Factors[1] (Table 31) (Decimals omitted). 138

Table 33 Rotated Oblique Third-Order Factor Coefficients and Correlations of Variables With Third Order Factors (Decimals omitted). 139

Table 34 Fourth-Order (General) Factor (Decimals omitted). 142

I. Statement of the Problem

The difference between basic and applied research, according to one definition, is that in applied research one seeks an answer to a definite question, whereas in basic research one is not even sure what the question is. By this definition research on individual differences (IDs) in human learning must be characterized generally as very basic indeed. The present study is no exception. A comprehensive review of the literature on this topic leaves a definite impression that at present we are still at the exploratory stage of trying to formulate potentially fruitful questions and devise adequate methodologies for answering them.

Though the state of the art is far from laudable, the small handful of researchers who have ventured into this untamed territory have not worked in vain and are deserving of credit for their efforts. We have learned something even from the seemingly most unrewarding attempts. The various pitfalls and blind alleys that they have brought to light might at least be regarded with caution in our future efforts. The total body of research on any problem, of course, always contains much more dross than gold, and this is especially true of the research on IDs in learning, the present study not excepted. But this must be more or less inevitable in the early stages of groping into uncultivated regions.

From an historical standpoint the meager progress of research on IDs in learning has been due, first, to the fact that experimental and differential psychology have developed from the beginning along quite different and separate paths - a theme elaborated upon in Cronbach's frequently cited Presidential Address on the "two disciplines of scientific psychology" (Cronbach, 1957) - and second, to the fact that the pioneering research on IDs in learning yielded such puny results, with little relevance at the time to the theories and problems of either experimental or differential psychology, as to discourage further pursuit of this line of inquiry. Moreover, none of the early investigators grappled with the problems long enough to really get them off the ground. There were no martyrs with the fortitude of an Ebbinghaus to give the necessary initial push. In fact, with the one exception of Fleishman's extensive and continuing investigations of IDs in perceptual-motor learning (see Fleishman, 1962, for a review), there are no other examples of long-term systematic programs of research on IDs in learning comparable to investigations of other important topics in psychology, such as Pavlov's work on conditioning, Skinner's on schedules of reinforcement, Thurstone's on the primary mental abilities, Eysenck's on extraversion-introversion, or Underwood's on distribution of practice in rote learning, to name but a few prominent examples.

Yet no one has gone on record as seriously questioning the importance of IDs in the scientific study of learning. It is quite generally agreed that we will eventually have to come to grips with this realm of phenomena if we are to achieve a comprehensive account of the psychology of learning. The key questions now concern how we should go about the business of giving the problem of IDs the attention it deserves in the experimental psychology of learning, where we should begin, what kinds of mistakes we should try to avoid, and what theoretical goals we should envisage.

A. General Background of the Present Study

The above questions were explicitly in mind in planning the present study. Though various approaches to these questions would be at least implicit in any historical review of the research on IDs in learning, it is not the purpose of this introductory statement to present a comprehensive review of this area. There already exists a number of general reviews of the major efforts in this area which, taken all together, provide a fairly adequate picture of the current state of our knowledge of IDs in learning (see Allison, 1960; MacKay & Vernon, 1963; McGeoch & Irion, 1952; Noble, 1961; Stake, 1961). The best historical summary is by Glaser (in press). Our purpose here is simply to outline some of the major types of problems and approaches that have arisen in research on IDs in learning and to indicate the general philosophy that underlies the present investigation.

1. The Basic Problem. - Let us approach the full complexity of the problem in several steps. First, imagine that we gave a large battery of various learning tasks to a large random sample of the general population. The simplest, if most unlikely, thing we could find would be that all the subjects (Ss) in our sample have the same score, within the margin of error of measurement, on every task in the battery. That is, if all scores on every task were converted to standard scores (z scores), both the mean and variance on every task would be zero. In short, there would be no problem of IDs at all. This would be a great blessing to the experimental psychology of learning, because it would mean, for one thing, that no more than one S would be needed in each of the various conditions of the independent variables in our experiments. The only reason anyone could have for using two or more Ss in each condition would be to reduce errors of measurement in the strict sense of the term. Many experimental studies of learning have been carried out as if this imaginary state of affairs was indeed the true state of nature, and this assumption is implicit in the conclusions drawn from many such studies. Unfortunately nature is not so ideally simple. Everybody knows that IDs exist. But do they really make any difference to the conclusions drawn from experiments which do not take IDs into account?

Suppose that in the sample of Ss that took our battery of learning tasks we do find true IDs in the Ss' scores. In other words, there would be some true variance among the scores on each of our tasks. But then suppose that we obtain the correlations between the scores on each task and every other task, and we find that the correlations are all perfect. In other words, all Ss would remain in the same rank order of ability on every kind of learning task. In this case, only one general factor, which we would call "general learning ability," would be needed to account for all IDs in learning. The hope of finding something approximating this state of affairs was the motivation of many of the early studies of IDs in learning. This hope was frustrated by every attempt and in the past decade researchers have more or less abandoned the pursuit of the factor of "general learning ability." Nearly every review of the literature seems to take the form of a chronicle of repeated failures along this overly optimistic path. To anyone who has been immersed in this literature, the statement that there is no factor of general learning ability must seem quite tiresome by now. But the situation is concisely summarized by Spies (1959), who reviewed all the factor analytic studies up to 1959. The picture has not changed since then. Spies wrote: "With one or two exceptions, no general factor has been found sufficient to account for the intercorrelations of scores on a number of different learning tasks. Garrett (1928) found a small general factor (tentatively identified as a memory factor) in a group of memory and learning tests by using Spearman's tetrad difference method. Langdon (1937) found a general factor to account for the intercorrelatins of scores on a group of learning tasks at each of four stages of practice. However, the general factor differed at each stage. Other than these studies, all investigators have found it necessary to use from 3 to 10 factors to account for their matrices of intercorrelations of scores."

Thus we must grant that different _kinds_ of learning tasks involve different ID factors. So let us look now at a single type of learning in hopes of finding some kind of simplicity with respect to IDs. We give our imaginary sample of _Ss_ two equivalent lists of paired associates to learn. One list is learned under conditons of massed practice, the other under distributed practice. We know what we would find, because variations of this experiment have already been done, first by Hovland (1938) and later by Madsen (1963. The correlation between the learning scores under massed and distributed practice turn out to be appreciably less than unity, so obviously more than one factor is involved in IDs even in paired-associate learning. Hovland (1938) found, for example, that some 44 percent of the _Ss_ in his experiment learned more rapidly under distributed than under massed practice, while some 38 percent learned more rapidly under massed than under distributed practice; about 18 percent of the _Ss_ showed no effect of the experimental variable one way or the other. Madsen (1963) found a complete reversal of the effects of distribution of practice as a function of measured intelligence. _Ss_ with low IQs benefited markedly from distributed practice; high IQ _Ss_ did better under massed than under distributed practice. So we see that not only the type of learning, but also the _conditions_ of learning, can be an important source of IDs.

Sensory modality of the input might also be another source of the IDs in learning, although this has not yet been demonstrated in normal _Ss_. In certain types of aphasia, however, learning can be seriously impaired when the material is presented, say, visually, and can be quite normal when presented aurally, which at least suggests the possibility of IDs in learning as a function of sensory channel.

A further type of complication possibly exists, though it has not yet been clearly demonstrated in the realm of learning. It is possible that two tasks of the same _type_ of learning (e.g., paired-associate learning, composed of the same _content_ (e.g., CVC nonsense syllables matched for frequency, association value, pronunciability, etc.), administered under the same _conditions_ (e.g., a 2-2 rate of presentation with massed trials), but differing in _quantity_ of material (e.g., 10-item vs. 20-item list) might also show an interaction with _Ss_, thus constituting one more source of IDs.

It is known in psychometric research that test items which appear to be extremely homogeneous and "unidimensional", but which differ in difficulty, can involve different factors The most striking demonstration of this fact is a study by Guilford (1941), in which intercorrelations were obtained among measures of simple pitch discrimination taken at ten levels of difficulty, ranging from differences of 1 to 30 cycles per second. Factor analysis of the intercorrelations revealed that three different factors are involved in making pitch discriminations over this range of difficulty. Analogous findings might well turn up in the realm of learning.

Still another source of IDs might arise from the stage of practice on a particular type of task. Ss might not remain in the same rank order of ability in learning, say, ten equivalent paired-associate lists. Ss may differ in rate of "learning-to-learn", or different abilities may come into play at different stages of practice, or both these conditions may exist simultaneously. Fleishman (1962) has shown that the factorial composition of certain motor learning tasks changes at different stages of practice. Whether this is true also of other types of learning is not yet known, but it seems highly likely.

2. Phenotypes and Genotypes. - To summarize what has been said, IDs in learning can arise from many different sources, each of which can be thought of as an interaction between Ss and some independent variable or set of many independent variables and their interactions. A few of the independent variables that most readily come to mind are: type of learning (e.g., conditioning, discrimination learning, selective trial-and-error learning, perceptual-motor learning, rote learning and its subcategories such as paired-associate and serial learning, free recall learning, conceptual learning, and so on); content of the learning materials (e.g., verbal, numerical, spatial, etc.); sensory modality of stimulus presentation (e.g., visual, auditory, kinaesthetic, haptic, etc.); and, finally, a host of experimental variables (e.g., rate of stimulus pacing, distribution of practice, stimulus duration, CS-UCS interval, degree of intratask similarity, meaningfulness, length of task or amount of material, difficulty level, instructional sets, meaningfulness, stage of practice, etc.).

Now, whenever the correlation between scores on any two learning tasks that differ on any one or a combination of the above variables is less than perfect (after correction for attenuation) we can claim to have found a phenotypic source of IDs in learning. As the term is used here, a phenotype is a significant interaction between Ss and any of the kinds of independent variables that have been listed above. Viewed in this way, one gets the impression that the number of phenotypes is apt to be practically overwhelming. Unfortunately, this is probably the correct impression.

If there were no significant correlations at all between learning scores under all these variations in the multitude of independent variables, the situation would indeed be hopeless. There would then seem to be no basis for developing a science of IDs in learning. We would never be able to progress beyond merely describing phenotypes. From a scientific standpoint this would seem to be a most unexciting and unrewarding enterprise. Therefore, we must begin with what seems to this investigator to be a wholly warranted article of faith, viz., that a limited number of basic genotypes underlie the welter of phenotypic IDs in learning. It is the major task of research on IDs in learning to discover what these genotypes are. Genotypes, in other words, are the main dimensions or factors in terms of which we can describe, and ultimately explain, the phenotypic variation of IDs in learning in its various forms under various conditions of learning.

What these genotypes will look like when we find them, if we ever do find them, is strictly conjectural at present. But the idea that they are there to be found seems worth pursuing. It seems most unlikely that learning abilities should be completely chaotic, without structure, and specific to every single task and condition of learning. If learning abilities are correlated with brain mechanisms and are a product of biological evolution, we should expect to be able to find some definite structure to IDs in learning. There is a system to every other type of biological functioning and it would be hard to see why brain and behavior should be exceptions. This point is emphasized because there are some psychologists who would discourage research on IDs on the grounds that IDs are such a mish-mash of untold sources of variance that we could never sort them out or make any scientifically satisfactory sense of them. The view held here is that the problem can be approached without complete pessimism or the risk of utter discouragement if one is willing to break the problem down into very small bits and proceed to tackle each of these one at a time.

To do this one needs to begin with some hunches or hypotheses about the kinds of genotypes one might expect to find These, of course, are dictated largely by our theoretical conceptions of the nature of learning and are bound to change as our knowledge increases and new theoretical formulations evolve. Some theoretical constructs have had considerable viability over a long period, however, and it is these which perhaps provide the best suggestions of where we might look for ID genotypes. Some of the phenomena associated with the following basic psychological processes and hypothetical constructs that we might look into in this connection are: (a) the rate of decay of stimulus or associative traces in the absence of rehearsal; (b) the rate of consolidation of stimulus traces; (c) the initial strength of reception of stimuli; (d) proneness to interference with trace consolidation or external inhibition of the trace; (e) degree of susceptibility of consolidated association to interference through associative competition; (f) reactive inhibition and stimulus satiation; (g) the degree of spread of primary and semantic stimulus generalization gradients; (h) drive, arousal, and attention; (i) "personal tempo"; (j) oscillation tendency, that is, the tendency for response strength to fluctuate up and down over the course of practice. Some of these factors may, or course, be derivatives of others. For example, oscillation tendency may be just one manifestation of interference effects, and associative competition might be based on differences in the breadth of generalization gradients.

B. Approaches to the Study of IDs in Learning

1. Intrinsic and Extrinsic IDs. - It will be evident that so far we have discussed IDs strictly in terms of variables and processes that are an inherent part of learning. These can be called intrinsic sources of IDs. They are IDs that exist in the processes of learning, and it is proposed that the genotypes underlying these IDs will probably have to be sought within the learning realm itself. On the other hand, we have what might be called extrinsic sources of IDs. These are subject variables which bear no resemblance to the learning process but which may influence, or be correlated with, performance in learning situations. The chapter on IDs in learning in the textbook by McGeoch and Irion (1952) is organized largely in terms of the research on extrinsic IDs in learning, probably because most investigators have taken this approach to the study of IDs. Performance on learning tasks is investigated in relation to such variables as chronological age, mental age, sex, personality traits, race, social class background, and other personal characteristics. The learning process is seen as being essentially the same for all individuals but as being influenced by all these factors such as age, IQ, personality, etc. These influences are what make for IDs in learning If we could select a group of Ss who were homogeneous on all these kinds of factors, according to this conception, we could eliminate IDs in learning. Implicit in this view is the idea that learning can be studied quite independently of IDs provided one takes adequate care to control for extrinsic IDs in one's experiments. It is true that these extrinsic IDs contribute to between-subjects variance in learning, but even when this variance is completely accounted for, a great deal of true variance remains. The sources of this remaining variance are probably intrinsic to the learning process.

From a theoretical standpoint the description of the effects of extrinsic IDs on various types and conditons of learning cannot in itself be seen as a satisfying goal. While these kinds of findings may often be of considerable importance to applied psychology, it must be admitted that they offer little to advance our understanding of the psychology of learning. Without knowing something about the ID genotypes intrinsic to learning, the facts of extrinsic IDs cannot be integrated with learning theory. Their modus operandi must remain a mystery in the absence of knowledge of the dimensions of IDs in the elemental processes of learning.

Not all personality traits are necessarily to be regarded as extrinsic to learning. In some cases there can be a theoretical link between the genotypic aspect of the personality trait and some intrinsic aspect of learning. A case in point is Eysenck's theory of extraversion-introversion, in which the underlying basis of this dimension of personality is said to be IDs in the rate of formation and dissipation of inhibition in the Pavlovian sense (Eysenck, 1957). Thus, in learning situations involving inhibition, one should expect to find differences between introverts and extroverts. But the phenotypic trait of extroversion also has associated with it certain habits and attitudes which may affect the person's performance in any learning situation, whether it involves inhibition or not. In this case we would be dealing with an extrinsic and more superficial aspect of IDs in learning. More extroverted Ss, for example, might be more careless about complying with the experimenter's instructions or might simply perform in a slap-dash way in order to get through a boring task as quickly as possible. Anxiety or neuroticism might also have intrinsic effects in learning, in so far as these traits act as drives or energizers of habit strength under certain conditions. In a sense, we are dealing in these cases with a single genotype which has behavioral manifestations both in the realm of personality and of learning.

2. The Psychometric Approach. - Here the aim has been to account for IDs in learning in terms of factors derived from psychometric "reference tests". Thurstone's tests of the Primary Mental Abilities (PMA), measuring such factors as verbal fluency, numerical ability, spatial ability, etc., are perhaps the prototypes of all such tests.

With minor variations, studies of this type have consisted of giving a battery of reference tests to a large number of Ss, along with an assortment of learning tasks, obtaining intercorrelations among the measures derived from all these procedures, and factors analyzing the whole lot by one method or another (e.g., Allison, 1960; Games, 1962; Stake, 1961).

A number of rather serious criticisms can be made of this approach, and these apply even to those studies which appear to be highly competent and sophisticated from a psychometric standpoint. The three fairly recent studies referred to above are all excellent examples. The major difficulties encountered in these, as well as in a number of earlier, less ambitious efforts (see Spies, 1959), are the following.

(a) Simply in terms of the logic of scientific explanation, it seems highly questionable whether one should even hope to be able to account for IDs in learning in terms of IDs on the kinds of status variables measured by psychometric reference tests. In terms of process, many of the learning tasks that have been used in this class of studies appear to be much simpler than any of the reference tests in terms of which the variance on the learning tasks is to be explained. From an analytical standpoint we have even less adequate understanding of the sources of variance on the reference tests than on the learning tests. Since most ability measures can be regarded as representing some point on the person's learning curve in the acquisition of some kind of knowledge or skill, it seems more reasonable to believe that IDs on psychometric tests will ultimately have to be "explained" in terms of more elemental learning processes rather than the other way around. Thus, to this investigator, the chief aim of the psychometric approach to IDs in learning seems rather futile at the very outset.

(b) When reference tests and learning tasks are factor analyzed together there always results a marked cleavage between the two types of measures. The aspects on which they usually show some overlap have to do with the content of the materials, such as reference tests and learning tasks which involve verbal or spatial abilities. The most obvious sign of this lack of communality between the two domains is the small amount of variance and covariance among the learning tasks accounted for by the reference factors. In the first place, the communalities of the learning measures are generally very low, which is a sure sign that the investigation is falling far short of realizing its ambitions. It means that most of what we are trying to capture has somehow eluded us. In most studies we do not know whether this was because the learning measures were too unreliable (since reliability sets the upper limit of the communality) or because the measures really do have very little variance in common. In the absence of reasonably high communalities, reliability estimates of the learning measures are essential if the results of the factor analysis are to be interpretable. This requirement has seldom been observed in factor analysis studies of IDs in learning.

The most likely reason for the cleavage between psychometric and learning tests is that the two classes of tasks have little in common in terms of procedural variables. Whatever ID variance in learning tasks results from interactions between Ss and such independent variables as pacing, distribution of practice, etc., is simply not tapped by the reference tests. The abilities measured by the reference tests may represent sources of transfer from previous learning to performance on the particular learning tasks being used in the factor analysis, but the process variables that play so important a role in learning and which probably constitute a major source of ID variance have little or no counterparts among the reference tests.

But this is only one of the possible reasons for the low communalities among the learning tests. Another reason is that most studies have used a large assortment of unrelated learning tasks which differ so greatly from one another in terms of procedural variables and the kinds of processes they involve that it would be most surprising to find substantial correlations among the tasks, unless, of course, one still believed in the existence of a large general factor of learning ability. When the learning tasks differ on nearly every conceivable dimension of content, difficulty, procedure, and all the other independent variables we mentioned earlier, substantial communalities and robust factors simply cannot emerge.

Furthermore, these studies have used for the most part learning tasks which have no previous history of having yielded any information about the nature of learning. The types of learning for which we have some background of experience from the experimental psychology of learning, such as conditioning, serial and paired-associate learning, etc., are conspicuous by their absence in factor analysis studies of learning, with the exception of the study by Games (1962). There would seem to be considerable value in having some knowledge of the main experimental parameters of the tasks we use in our studies of IDs. We can then enter different values of these parameters into our procedures and get whatever ID variance is associated with them into our factor analysis.

The writer can give an example from his own experience of the importance of these procedural variables. A significant correlation of .45 was obtained between serial learning and college entrance examination scores (ACE Test). The serial learning had been paced at a 4-sec. rate and was therefore fairly easy for college students. It was thought that by making the task more difficult, thereby obtaining a greater spread of scores, it would be possible to considerably boost the correlation between serial learning and ACE scores. So the serial learning was administered to another group of Ss, this time at a 2-sec. rate of pacing. Under these conditions the correlation between ACE scores and serial learning almost completely disappeared and was not significantly greater than zero (with N = 60). The variance was much greater for the 2-sec. rate than for the 4-sec. rate, and from this fact alone one would have expected a higher correlation for the 2-sec. rate. What had happened? Fortunately the Maudsley Personality Inventory was available on all the Ss in both groups. It was noted that the Neuroticism scale was not significantly correlated with serial learning ability at the 4-sec. rate but was significantly correlated at the 2-sec. rate. When the task was made more difficult and stressful, the Neuroticism factor tended to override intelligence in determining IDs in performance on the serial learning task. Thus it can be seen that a single procedural variable, in this case pacing rate, can markedly affect the pattern of intercorrelations among scores on a learning task and psychometric and personality measures. Similar findings have been reported by Eysenck (1965) in the case of eyeblink and GSR conditioning.

(c) The psychological descriptions of the factors that emerge from factor analytic studies of learning generally consist of the names of well-known categories of learning tasks, or of the content of the learning materials, or of some combination of the two. From the standpoint of any theory of learning, these kinds of factors hardly seem enlightening. The factors involving learning in Stake's (1961) study were labeled "rote memory," "memory-task learning," and "numerical-task learning." The principal learning factors in Allison's (1960) study were "verbal conceptual learning", "mechanical-motor learning," "spatial conceptual learning," and "rote learning."

One methodological aspect of Stake's (1961) study is perhaps deserving of comment, not because it makes much difference to the kinds of factors that were bound to emerge from his data, but because it probably blurred the factor structure considerably. Stake included both Negroes and whites in his study and put this dichotomy into his factor analysis as a "Race" variable. The first factor was rotated to "Race" and thereby absorbed whatever variance was associated with this variable, which in this particular case happened to be a rather large general factor. The communality of Race in Stake's battery of tests was .81. The residual intercorrelation matrix after Race was partialed out indicates that there were pretty slim pickings for the determination of the remaining factors. Using Ss who span a wide age range and including age as a variable in the factor analysis is apt to have similar results. By including extrinsic, nonpsychological variables such as race and age in our factor analyses, we run the risk of burying a considerable proportion of the variance in factors which permit no psychologically enlightening interpretation. The "factor" of Race, which sopped up so much of the variance that might have been psychologically more enlightening if it had been distributed among other factors in Stake's analysis, is a good example of this.

C. Learning Ability and Intelligence

One of the major empirical and theoretical tasks of research on IDs in learning is the integration of this domain of phenomena with the domain of psychometrically measured abilities. One aspect of this general problem - the relationship between learning ability and measured intelligence - has been a more frequent topic of research in this field than any other. Early studies, notably those of Woodrow (e.g., 1946), gave rather widespread currency to the notion that, to use Woodrow's words, "The ability to learn cannot be identified with the ability known as intelligence." This conclusion was unacceptable on commonsense grounds (which is, of course, not a scientifically valid reason for rejecting it). Many other researchers have pursued the problem up to the present time, with the result that much of the work of Woodrow and other early investigators has been found to be handicapped by serious methodological shortcomings which tend to invalidate their conclusions. The consensus of more recent research is that there is a substantial relationship between learning ability and measured intelligence and that the exact magnitude of this relationship depends upon the independent variables operating in the learning situation. Some independent variables interact with traits other than those measured by intelligence tests and when such interactions are powerful they can override the correlation between measured intelligence and performance on learning tasks. Because learning is itself not a unidimensional phenomenon, it should not be surprising that it does not always maintain the same degree of correlation with intelligence. From this standpoint it is almost surprising to see the large variety of conditons under which many kinds of learning do, in fact, show a substantial relationship to intelligence. The literature on this topic has been thoroughly reviewed by Rapier (1962), and Jensen (1965) has more recently reviewed the evidence with respect to serial and paired-associate learning.

The view taken here is that the correlation between learning abilities and what we measure by means of intelligence tests is, in a causal sense, a rather indirect relationship, resulting from the fact that intelligence tests for the most part measure the amount of knowledge (or attentional, perceptual, and verbal and symbolic habits and skills) acquired through cumulated experiences up to a given age in what is assumed to be a more or less average cultural environment. Thus, indirectly this index of the amount of acquired knowledge reflects the average rate of acquisition under almost infinite variations of the conditions of learning. Intelligence might also be determined in part by IDs in the availability and generalizability of previous learning for transfer to relatively novel problems. These factors would show up in the learning domain as well as in performance on intelligence tests of the problem-solving type. A similar view of the relationship between learning and intelligence has been elaborated upon by Ferguson (1954, 1956). The point being made here is that a more basic understanding of the nature of "intelligence" can be gained through the study of IDs in learning processes. Learning abilities are not seen as something to be explained in terms of "intelligence." The reverse would seem to be a psychologically more defensible position.

1. Gain Scores. - The main methodological weakness of early studies in this area was their use of so-called "gain scores" to measure learning. Raw gain is obtained by subtracting level of initial performance from level of performance after a specified period of practice. There would be no problem with this measure if initial performance were the same for all Ss, in which case one would not need to use gain scores at all, since it would only amount to subtracting a constant from everyone's final performance level. The use of raw gain has often resulted in such artificial findings as negative correlations between learning (as measured by gain) and IQ. The various hazards of using gain scores to measure learning and the possible satisfactory solutions to this problem have been thoroughly explicated by Manning and Du Bois (1962). They propose the use of "residual gain" scores. Residual gain consists, in effect, of some measure of final level of performance from which initial performance has been partialed out by statistical regression. In other words, final performance is divided into two parts, a part predictable from initial performance level and a part which is unpredictable from initial performance level. Putting initial and final measures into a principal components analysis or a factor analysis accomplishes the same purpose as the use of residual gain scores; if learning scores contain a source of ID variance not contained in the initial scores, this will show up in the factor analysis - the two kinds of scores will have different patterns of factor loadings In the present study both learning (incremental) scores and interference (decremental) scores are dealt with in this fashion.

D. The Definition of Learning

We are here concerned only with the operational definition of learning rather than with a theoretical definition. Learning is often defined operationally as a "relatively permanent" change in behavior resulting from experience The writer can see no justifiable reason for putting a time limit on the duration of the behavioral change for it to qualify as learning. It would be quite artibrary what the minimal duration of the change should be in order to call it learning, so it seems preferable not to limit the definition in this manner. By defining learning too narrowly we might exclude many important and relevant phenomena from the field of research on IDs in learning. Thus, to allow the widest possible scope to investigations in this field, the writer prefers to define learning as follows: Learning is said to have occurred when there is a change in the probability of some specified response, or class of responses, following the cessation of a specified stimulus situation, excluding extrinsic physical or chemical alteration of the nervous system and changes due to physical growth. Nothing is said about the interval between the cessation of the stimulus and the response, the duration of the stimulus, or the length of time through which the change in response probability must be maintained. To take an extreme case, even reporting correctly what one sees on a very rapid tachistoscopic exposure would be regarded as a form of learning. Dialing a telephone number correctly two seconds after reading the number from the directory would be another instance of learning. Some stimulus situations result in changes in response probability that last only a matter of seconds while others may last a lifetime. The relationship between phenomena at various points along this continuum is a matter for empirical investigation. There seems to be no good reason why any segment of the continuum should be ruled out of the learning domain by definition.

E. Innate and Environmental Sources of IDs in Learning.

IDs in learning as measured in the laboratory are always a mixture of innate dispositions and transfer from previous learning experiences. Beyond the first weeks of life it is probably impossible to observe innate IDs in learning in a "pure" form. However, different forms and conditions of learning are undoubtedly influenced to varying degrees by innate, constitutional factors, on the one hand, and by environmental and historical factors, on the other. Both sources of IDs are important, of course, and neither should be neglected in research on IDs in learning. The fact that we are dealing with two main sources of variance, however, constitutes a problem in the conduct of our research. The question is how to deal with it. One solution is to attempt to investigate these two sources of IDs more or less independently, at least at first. This would be done by carefully selecting learning tasks for their suitability in one type of study or the other. Learning tasks would be selected, by means of task analysis based on our psychological judgment, for inclusion in one of two classes, one consisting of tasks that seem to depend very little on transfer from previous experience, and the other consisting of tasks that clearly involve previous learning. It seems advisable in the early stages of our research to keep these two classes of tasks as separate as possible, otherwise there is the danger that one source of IDs will get in the way of our attempts to sort out the structure of the other source. Environmentally determined IDs will be easier to investigate if we can select and match subjects in terms of the relatively innate types of IDs. IDs in the process variables are more apt to be of the "innate" variety, while IDs of the "environmental" variety are more apt to be reflected by content variables. For example, we might expect learning or retention of a letter series to have a greater "environmental" loading than a digit series, due in part to the fact that individuals have previously acquired quite definite letter sequence habits, while digit sequence habits, other than ordinal numeration, are relatively equal for all random combinations of digits; linguistic background might therefore interact more strongly with performance on letter series than on digit series. Learning in most "real life" situations is generally highly dependent upon transfer from previous experience, and for practical purposes an assessment of a person's past experience may be more important than a knowledge of his basic learning abilities for predicting his performance in some kinds of learning situation.

F. The Approach of the Present Study

The present study is concerned with what is presumed to be one of the basic sources of IDs in various learning phenomena, viz., susceptibility to interference effects. Interference tendency is hypothesized as one of the genotypes of IDs in learning. We wish to find out (a) if IDs in the effects of interference are psychologically distinguishable from IDs in learning and (b) if IDs in phenotypically different kinds of interference effects are referable to a common genotype.

II. Objectives and Design of the Study

This study has two main objectives: (a) to devise a number of laboratory measures of IDs in susceptibility to interference effects in learning, and (b) to determine the dimensionality of IDs in a variety of phenotypically different kinds of interference effects.

The study was organized into two main divisions corresponding to the two objectives stated above. The first part of the study was concerned with the selection, development, and assessment of learning tasks and procedures for measuring IDs in interference. The second part of the study consisted of a factor analysis of all of the measures finally selected. The chief purpose of the factor analysis was to determine whether the phenotypically different kinds of interference effects represent a single common factor (i.e. the same genotype) or a number of different, though possibly correlated, factors. Practically all learning tasks are expected to have some loading on interference of one kind or another. Some tasks, however, should be more highly loaded with interference than others. Factor analysis should reveal whether our experimental tasks load on interference factors proportionaltely in accord with our hypotheses concerning the amount of interference involved in the task.

A. Underlying Philosophy of this Research

At least seven more or less implicit hypotheses influenced the general plan of this investigation:

a. The domain of learning tasks should be fairly limited. This means that in the initial stages of investigating IDs in learning, we should not throw too many widely diverse types of learning tasks, or tasks haphazardly selected from all the various categories of learning, into one hopper for factor analysis. If we begin with too much diversity, the proportion of task-sepcific variance is apt to be so great as to attentuate and obfuscate the structure of genotypic IDs. Many of the earlier studies were more or less hopelessly vitiated for this reason. On the other hand, the task domain should not be so restricted as to risk precluding the emergence of any factorial structure at all. The extreme case would be that in which all variability was due to a single general factor plus error variance. The range and types of learning tasks that can be profitably used in any single factor analytic study of IDs in learning at the present stage of our knowledge are at worst a matter of trial-and-error and at best a matter of slightly informed guesswork. (The situation in the case of motor learning is already considerably better off in this respect than other areas of learning, due largely to the extensive research of Fleishman (e.g. Fleishman, 1962) on IDs in perceptual-motor skill learning.)

b. *Begin by trying to find some one, basic genotype of IDs in learning.* This means that one should not merely throw a number of tests into the hopper in hopes that factor analysis itself will do all the work in pointing up interpretable factors or genotypes. It is probably much better to begin by hypothesizing the existence of some ID genotype and then to devise tests that would seem to tap this genotype to varying degrees. Determining the existence of the hypothesized factor only implies being able to distinguish it from other sources of ID variance in the task domain under investigation; it does not necessarily require that all the other independent factors that emerge from the analysis need be hypothesized beforehand or precisely identified after the analysis, though surely one would hope that they would give rise to further hypotheses that could be specifically investigated in subsequent studies. More often than not the hypothesized factor that one sets out to "capture" in a study especially designed for this purpose will probably turn out to consist, not of one, but of two or more factors. Further experiments will usually be necessary to substantiate and clarify these originally unforseen factors. But in the early stages of this research the investigator will have problems enough if he sets out to capture a single hypothesized factor and uses a variety of tests which are presumed to be heavily loaded with the factor along with "background" tests which are presumed to be minimally loaded on the hypothesized factor.

b. *The learning tasks should be analytically simple and should have high phenotypic "face validity" for the hypothesized factors.* Analytic simplicity, as the term is used here, means that the input and output, or stimulus and response, aspects of the learning task be clearly specifiable and quantifiable. The more completely and clearly that we understand the independent variables that affect the dependent variables in our learning tasks, the better off we will be. This implies two things: reliance on an experimental (i.e., manipulating independent variables) approach, and reliance, at least in the early stages of investigation, on traditional and classical learning tasks and paradigms about which we already have a substantial background of information concerning relevant independent variables, etc. Failure to observe this principle has led to trouble in past investigations of IDs. The odd assortment of "learning" tasks used by Woodrow (1938, 1946) in his unenlightening and misleading studies of IDs in learning are a good example of how the use of tasks which have no previous history of having yielded fruitful knowledge about the learning process can result in meager and puzzling results when used to study IDs.

From the beginning the present study attempted to maintain analytic simplicity in the choice of learning tasks, but unfortunately the principle of using already thoroughly tried and true types of learning tasks was not observed until after a number of initial failures with newly devised learning tasks. These tasks were devised mainly with the intention of achieving a high degree of control over the stimulus and response and reinforcement elements of the learning task. Most of the tasks consisted of trial-and-error selective learning, in which the subject had to learn S-R associations between stimuli displayed on a screen and a number of pushbuttons on a "response console." The reinforcement for correct responses was a green light. At face value the tasks had great analytical simplicity and the interference conditions that could be "built" into these tasks had great face validity. In brief, viewed from the theoretical armchair the tasks seemed ideal for the investigation of IDs in susceptibility to interference effects in learning. The only trouble was that these seemingly neat tasks did not work. It proved impossible, for all practical purposes, to achieve satisfactorily high test-retest reliability of measurement with these S-R tasks. The reason seemed to be that these tasks were actually much less simple in terms of the controlling variables than anyone had originally imagined. (The details of these problems are discussed in the section on Procedures.) Other tasks patterned after more traditional measures of ability and learning (memory span and serial learning) turned out much more successfully.

_d_. _Process and procedural variables are more apt to reflect basic and pervasive genotypes of IDs in learning than content variables_. This is only a hypothesis, which, of course, will have to be tested in future research. But one has to begin somewhere, and it is the investigator's hunch, based on an examination of past research of IDs in learning, that the largest and most general sources of ID variance in learning tasks are associated with the conditions of learning rather than with the content or substantive aspect of the learning material. Reference tests constructed along the lines of content variables (e.g. verbal, numerical, spatial, etc.) generally account for only a small proportion of the variance among subjects performing on learning tasks whose content is intended to represent these reference factors (e.g. Allison, 1960; Stake, 1961). The reason may well be that most of the ID variance arises from the Subjects X Procedural Variable interaction. Psychometric reference tests would not be expected to tap this source of variance in learning.

It also seems reasonable to hypothesize that content sources of variance are more apt to refelct transfer from previous learning and in this sense would represent less basic, or more phenotypic, sources of ID variance than that which arises from process and procedural variables. In traditional laboratory learning procesures this class of variables is represented by such familiar terms as distribution of practice, intratask similarity, stimulus and response pacing reates, proactive and retroactive interference paradigms, the length-difficulty relationship, and so forth. Another reason for believing that more fundamental psychological riches are likely to be found in this area is that it seems easier to conceive of the biological evolution of the brain, and hence of its most basic learning mechanisms, in terms of such fundamental aspects of oragnismic-environmental interaction as those involving time relationships among stimulus variables, intensity and similarity gradients among stimuli, biological salience of the stimuli, work requirements of responding, and so on. These factors are reflected in some of the most fundamental constructs that have emerged from the experimental psychology of learning, such as stimulus generalization, the gradient of reinforcement, work inhibition, stimulus satiation, proactive and retroactive interference, drive, neural consolidation, and so on.

e. Psychometric reference tests are not likely to clarify the dimensionality of learning processes. This is partly a corollary of the preceding assumption. There is little if any evidence that the processes of fundamental interest in learning are tapped to any appreciable extent by the usual paper-and-pencil psychometric tests of abilities. What these tests get at is probably best viewed as a highly derivative amalgam of IDs in a host of basic learning processes and the multifarious past learning experiences of the subjects. The learning process variance in these kinds of tests, therefore, is going to be anything but pure or simple. It has always seemed ridiculous to this investigator to try to explain IDs in relatively simple and highly controlled laboratory learning tasks in terms of IDs in extremely complex and relatively uncontrolled psychometric tests. It amounts to the self-deception of explaining one unknown in terms of an even greater unknown. A more sensible way of proceeding would be first to work out the dimensionality of learning variables strictly within a particular learning domain (such as conditioning or short-term memory or rote learning) before trying to relate the learning factors to phenotypically very different classes of performance variables. Entering a large battery of psychometric reference tests into the same factor analysis with a number of learning measures usually has resulted in three rather frustrating consequences (e.g., Allison, 1960; Games, 1962; Stake, 1961, viz., (a) little of the variance in the learning tasks is accounted for or "explained" by the reference tests; (b) practically all the interpretable structure that emerges from the analysis is due to the reference tests, while the learning tests stand apart by themselves among a few emaciated, scarcely interpretable factors; (c) factors substantially resembling any of the basic processes that are central in the experimental psychology of learning almost never emerge, either because the initial selection of learning variables excludes such factors or because the factorial structure imposed by the reference tests cuts across and obscures the learning process variables.

f. Learning tasks on which there are likely to be large IDs in transfer from previous learning are best avoided in the pursuit of the more basic genotypes of IDs in learning. While transfer of skills from previous learning can never be entirely eliminated from laboratory learning tasks, it is surely possible to minimize the differential effects of transfer. Many traditional laboratory learning tasks were, in fact, originally devised for the purpose of minimizing extra-experimental transfer. There are two general means of achieving this: (a) by selecting tasks which have very few stimulus or response elements in common with the kinds of things subjects learn in their natural environment, or (b) by selecting tasks in which the basic elements are so much a part of common experience that virtually all subjects have reached practically the same asymptote of degree of learning of these elements.

This recommendation does not imply that transfer is an unimportant source of ID variance in many learning situations. It is certainly highly important in learning the usual school subjects. In fact, it is doubtful if IDs in any form of complex learning can be understood without reference to differential transfer from previous types of learning (see Gagné, 1965). But differential transfer effects may be conceived of as a different major source of IDs in learning than the more "primitive" IDs we are concerned with in the present investigation. More rapid progress in research is apt to take place if some effort is made to avoid unnecessarily confounding variance due to IDs in basic abilities or processes and variance due to differential transfer from the varied previous learning histories of the subjects.

g. The dimenstionality of IDs in learning will be established only through a long series of successive approximations. One cannot realistically imagine clearing up the field of IDs in learning by means of a few deft, large-scale factor analyses of learning variables. Factor analysis is viewed in this endeavor as only one of the means of discerning the major sources of IDs in learning. Experimental analysis and refinement of procedures is another means. But the application of these methods will not lead from chaos to order in one or two bold steps. It is expected rather that progress will be made, at least initially, by rather small increments resulting from a repeated cycle of procedures that will involve (a) hypothesizing a genotypic source of IDs, (b) devising tests to measure the hypothesized factor on the phenotypic level, (c) factor analyzing these tests along with other tests in the same domain but which are presumed not to be highly loaded on the particular hypothesized factor, (d) formulation of new hypotheses, based on the results of the factor analysis, concerning the nature of the originally hypothesized factors, (e) selecting the most promising old tests and devising new ones in accord with the newly gained insight into the nature of the factors in order to test these insights in another factor analysis. Also, different groups of subjects selected on the basis of their standing on a particular factor can be compared in an experimental learning situation in which the factor is hypothesized to be an important source of variability. When the factors operating within a limited domain are fairly well-established in this fashion, their presence can be sought in other areas of learning. As the sphere of inquiry is broadened, of course, additional factors will be encountered, and these then will have to be studied in their own right.

B. Test Selection and Development

The problem of test selection and development is of crucial importance at this stage of research on IDs in learning. In expenditure of time and effort it constituted by far the largest part of the present study.

Since we are looking for a factor of susceptibility to interference effects in learning, hereafter to be referred to simply as an interference factor, the basic plan of the study was to devise a number of pairs of tasks of which the two parts are phenotypically as much alike as possible, except that one task in each pair would be so devised as to be much more heavily "loaded" on interference than the other task in the pair. The two types of tasks are referred to as High and Low interference tasks, respectively.

If interference is not a factor independent of learning ability, there should be no significant interaction between subjects and the High _vs_. Low interference conditions of the tasks.

There are a number of phenotypically different kinds of interference that have been of interest in the experimental psychology of learning. A number of these phenotypically different types of interference were included among our tasks in order to determine whether they have a common genotype. Interference phenomena can be classified in a number of ways. In the present study the two main dimensions of task classification are (_a_) tasks involving "short-term" _versus_ tasks involving "long-term" memory, and (_b_) tasks involving intratask interference _versus_ tasks involving intertask interference.

Measures of interference in short-term learning and retention were based on memory span tests in which the subject was required to recall a seqquence of items more or less immediately after a single presentation. The long-term tasks consisted of serial rote learning by the usual method of anticipation, in which the subject has repeated exposure to the materials until they are completely learned. Such a rote learning task, of course, has aspects of both short and long-term retention, but it is presumed that repeated trials on a series of items allows time for consolidation of traces in the short-term store or in some way results in the items' being coded, by the end of learning, in a somewhat different form than that in which they were held in short-term memory.

The phenotypically different types of interference included among the short-term memory span tasks were (_a_) interference due to high intralist similarity, (_b_) proactive interlist interference, (_c_) retroactive interlist interference with high interlist similarity, (_d_) retroactive interference with low similarity between the series to be recalled and the interpolated activity, (_e_) interference created by requiring the subject to try to recall lists that greatly exceed his memory span, (_f_) interference due to being required to recall a series in the reverse of its order of presentation. The types of interference included among the serial learning tasks were (_a_) interference due to high intralist similarity, and (_b_) retroactive interlist interference.

Another interference task that cannot be classed under any of the above headings but was used in the present study is the Stroop color-word test (Stroop, 1935). No learning to speak of is involved in this test, which seems to be a pure measure of response competition. Response competition, that is, the tendency for strong habits or associations to block or interfere with weaker associations, is thought to be one of the mechanisms causing interference in learning and retention. It is of interest to know whether the kind of interference measured by the Stroop technique, which is quite obviously a form of response competition, has anything in common with the interference that occurs in learning.

1. Rationale of Test Selection. - Why were memory span and serial learning tests selected from among all the possible types of learning tasks? Indeed, one can question whether memory span has anything at all to do with learning. In the beginning, it was decided to use, in addition to the memory span and serial learning tasks, a number of tasks involving trial-and-error selective learning of stimulus-response (S-R) associations. For reasons discussed in the section on Procedures these selective learning tasks proved unsatisfactory for our present purposes, and a nujmber of variations and refinements of the memory span and serial tasks were added to take their place.

How can the inclusion of memory span tests be justified in a study of IDs in learning? Several good arguments can be made in favor of their use at this stage of investigation. In the first place, a memory span test is in many ways equivalent to the first trial of many kinds of traditional learning tasks. The resemblance between a single trial on a serial learning task and a memory span test is especially obvious. The memory span test is a measure of how much the subject retains after a single, short-term exposure of some material. A serial or paired-associate learning test is a measure of how much the subject retains after a given number of repeated exposures tothe material or, conversely, the number of repetitions needed to achieve complete recall of the material. Memory span is the amount learned in a single trial. Traditional learning procedures measure the number of trials (or total time) needed to learn a specified amount of material. At this level of analysis the difference between learning and memory span is merely formal rather than psychological.

To determine whether repetition introduces another factor to whatever the factors are that enter into memory span, one of the tests in the present experiment consisted of measuring memory span after two presentations of the same digit series. A single repetition of the same series does indeed produce a substantial increment in the amount recalled, and this increment must be called learning if learning is defined as improvement resulting from repetition or practice. The question, however, is whether this learning increment involves any different factors than those contained in the single memory span test. This is one of the many questions this study attempts to answer. It is an empirical question, of course, whether memory span and learning tasks have anything in common in terms of IDs, and, as we shall see, memory span and serial learning have much in common. The point being made at the moment, however, is that even at the level of formal analysis memory span tests are part of a continuum of rote-learning tasks.

Another practical virtue of memory span tests is that a great many of them can be administered in a single one-hour laboratory session, which means that the reliability of the measures obtained from the tests can be made very high, as one would expect from the Spearman-Brown formula relating reliability to the number of items in the test.

But the most important advantages of using memory span tasks in investigating IDs are twofold: (a) not only does memory span have a good deal of formal similarity with learning, but we know that memory span also has highly substantial correlations with a host of psychometric measures of intellectual ability; and (b) memory span tests, unlike most psychometric tests, readily lend themselves to the experimental manipulation of most of the experimental variables of major importance in the study of learning. Thus memory span can serve as a link between psychometrically measured abilities on the one hand and factors in learning on the other. One of the major goals of differential psychology is, of course, to integrate the structure of mental abilities as derived from psychometric tests with the structure of learning abilities. Furthermore, in studying memory span experimentally, we are at least assured that the subject matter of our inquiry is not just an isolated laboratory phenomenon but is highly correlated with other intellectual abilities which are of considerable consequence to every individual's performance in his everyday life. A thorough analytical understanding of the nature of memory span would probably constitute a major step toward understanding the psychological conglomeration called intelligence.

The relation of memory span to general intelligence has been uncritically and erroneously depreciated, and the misconceptions involved therein should now be dispelled. Memory span for digits formed a part of the original Binet intelligence scale and has been included in all the revisions of the test. Digit span is also one of the subtests of the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Intelligence Scale for Children (WISC). Wechsler, however, has stated that "...as a test of general intelligence it (digit span) is among the poorest" (Wechsler, 1958, p. 70). But let us look at the facts of the matter taken from Wechsler's own massive normative data on his intelligence tests. First of all, we must note the reliability of the digit span (DS) test of the WAIS is reported as between .66 and .71 for different age groups (Wechsler, 1958). The WISC Manual reports DS reliabilities between .50 and .60 for various age groups (Wechsler, 1949). The reliability of the Full Scale IQ on both these tests is between .92 and .97. Vocabulary has the highest reliability (.95) of any of the single scales of the WAIS. The raw correlation between DS and Full Scale IQ in the WISC, after correction for attenuation, ranges between .64 and .70. These correlations compare very favorably with those of the other individual scales after they are corrected for attenuation. Thus the depreciation of digit span as a test of intelligence seems to have been based, at least in Wechsler's case, on the failure to take account of its low reliability as compared with the much higher reliabilities of some of the other scales.

But it is also of interest that Wechsler has claimed that DS contains very little "g" or the general intelligence factor derived from factor analysis. Yet in Wechsler's book (1958) a factor analysis (Holzinger's bi-factor method) of the WAIS scales is presented in which a large "g" factor was extracted. The DS test had a loading of .63 on "g" in the age group 18 to 19, which is the peak age for memory span. Corrected for attenuation, this factor loading becomes approximately .80. Thus, even according to Wechsler's own data, Digit Span is very substantially loaded on the general intelligence factor common to all the tests in the WAIS battery. It is also of interest that the loading of DS on "g" gradually declined with age, becoming as low as .48 (uncorrected) in the age group of 60 to 75+. Tests that measure past learning or "crystalized" intelligence, such as Vocabulary and Information, on the other hand, show no appreciable change in loading on "g" with increasing age.

It seems reasonable to suppose that DS might correlate even more highly with Full Scale IQ if it were not for the fact that many of the other scales seem to be rather highly saturated with a cultural-educational factor, while memory span very probably is not. The ability to answer "Who wrote Faust?" would certainly seem to be more a function of environmental influences than the ability to recall a series of digits. And interestingly enough, Digit Span has the lowest correlation with amount of formal education of any of the WAIS scales. It is probably the most "culture fair" scale of the lot. This possibility alone should make the further development and elaboration of memory span tests as measures of general intelligence a worthwhile undertaking.

Another interesting fact is that Digit Span is the only subtest of the WAIS on which there is absolutely no difference between the sexes. This, too, is probably evidence of the memory span's independence from environmental influences. Gates and Taylor (1925) found, for example, that even though memory span could be improved in children by prolonged daily practice on digit span tests, the improvement was relatively short-lived, and after several months without practice the memory span was right back to the norm for the children's chronological age.

The Digit Span score on all the Wechsler tests consists of the longest series the subject can recall in the forward direction plus the longest series that can be recalled in the backward direction. One span might well be more highly related to general intelligence than the other, and adding the two together would only conceal this finding. The likelihood that forward and backward digit span tap different factors is suggested, for example, by the fact pointed out by Wechsler (1958) that patients with Korsakoff's syndrome have a much greater discrepancy between forward and backward span than normal persons. (Korsakoff patients have a relatively much poorer backward span.)

The first and only factor analytic study of memory span, by Brener (1940), resulted in information that further recommends the measurement of memory span as a technique for getting at quite basic and pervasive genotypes of intelligence. Brener's purpose was to determine the constancy and the factorial composition of memory span under variations of the type of stimulus materials and the sensory avenue of presentation (visual and auditory). He used digits, consonants, words, geometric designs, and colors. A large general factor was found and three other small, rather nondescript factors, only one of which is at all interpretable, being a verbal factor associated with consonants and words. Digits were loaded .86 on the general factor. The factor loadings did not differ appreciably for visual and auditory presentation. Thus memory span seems to have a good deal of generality and is not much influenced by content or sensory modality of the input. There are also indications that digits are a more suitable medium than letters for measuring memory span without introducing extraneous factors. There are fewer differences in sequential dependencies among digits than letters; letter combinations have higher association values than digit series; and acoustic confusions are more prevalent among letters than among digits. In short, memory span for digits would seem to be less influenced by perceptual abilities and by transfer from previously acquired associations than memory span for words or letters.

The choice of serial learning for inclusion in this battery of tests was based on at least three considerations. First of all, serial learning, more than any other standard rote-learning task, formally appears to be the closest analogue of short-term memory span tasks, and thus it seems reasonable to hypothesize psychological processes common to both serial (long-term) learning and memory span (short-term learning). Secondly, certain properties of serial learning have been found to hold true under a wide range of experimental conditions. Characteristics of serial learning such as the shape of the serial position curve, the proportion of correct anticipations to total opportunities, the greater skewness of the serial position curve for omissions than for overt errors, and the perfect linearity of the Vincentized learning curve are all practically constant for a wide variety of materials, pacing rates, intertrial intervals, and other experimental conditions. Almost the only feature of serial learning performance which seems to be experimentally manipulable is the speed of learning. Finally, serial learning appears to be factorially constant throughout the course of learning. That is to say, whatever ability factors or processes are operating in serial learning at any one stage of learning are the same at other stages of learning. This is not the case in all types of learning. In motor learning, for example, it has been found that different ability factors come into play in different amounts at various stages of practice (e.g. Fleishman & Hemple, 1956). A serial list, on the other hand, can be likened to a psychometric test which is factorially homogeneous in the sense of having very high interitem correlations. A serial list, in fact, is practically a perfect Guttman Scale. As we shall see in the discussion of the reliability of serial learning, the interitem correlations in serial learning are in the region of the interitem correlations characteristic of highly homogeneous or unidimensional psychometric tests. The advantages of this feature of a learning task used in the study of IDs seems obvious. A total score taken from the S's performance, such as number of trials to criterion or number of correct responses in x number of trials, is representative of the S's performance as a whole as well as of his performance on every item of the list at every stage of learning. Thus, in terms of measuring the speed of learning, serial learning is relatively simple and clear-cut, having the major advantages of a highly refined psychometric test, yet obviously being a measure of learning. Whether these features hold true for paired-associated learning is not known.

The inclusion of paired-associate tasks in the present battery would have been interesting but also would have undoubtedly complicated the factor analysis. There is reason to believe that paired-associate learning is not highly correlated with serial learning, at least in our college population, and that different processes are involved in the two types of learning (Jensen, 1962). To clarify the relationship between serial and paired-associate learning would call for a major research project in its own right. This investigator feels that paired-associate learning is psychologically much more complex than serial learning. Thurstone was the first investigator to include both types of tasks in the same factor analysis, as part of his well-known factorial studies of intelligence. He concluded that "...the memorizing of temporal sequence, as in digit span, letter span, Knox cube, and serial learning, involves a retentive ability that is different from the rote memorizing of paired associates" (Thurstone & Thurstone, 1941, p. 6.) If Thurstone's conclusion is valid, it poses an important problem for the experimental psychology of rote learning: How and why do paired-associate and serial learning differ? An answer to this question is long overdue.

2. Establishing the Hypothesized Effect of the Experimental Variable. - Once the tests have been selected, it must then be shown that the experimental procedures intended to produce the effect under investigation do in fact produce the effect, which in the present case is interference. Unless the phenotypic effect is operationally established, one cannot claim to be measuring IDs in this phenotype. Therefore, for each of the measures intended to reflect some type of interference effect, evidence must be presented that there is a significant performance decrement under the High Interference condition. In other words, to take one example, one could not claim to be measuring IDs in retroactive interference (RI) unless it were also demonstrated that the experimental conditions had in fact produced RI.

3. Establishing Valid Individual Differences in the Hypothesized Effect. - It next becomes necessary to rule out spurious IDs in the hypothesized effect. In the simplest terms, the critical distinction between authentic IDs and spurious IDs is that the latter can be eliminated by some monotonic transformation of the scale of measurement. For example, consider the case of learning under conditions of high (H) or low (L) interference. If there is a difference between the speeds of learning under L and H, we can say that our experimental variable that distinguishes the H and L conditions has had an effect on performance. But it does not necessarily mean that the experimental variable (e.g. some procedure producing interference) is a source of IDs.

There are three basic types of spurious IDs: the additive, the multiplicative, and the exponential. In the simplest, additive case $H = L+k$. Consequently we would find that the correlation between H and L is unity (corrected for attenuation) and analysis of variance would yield a nonsignificant Ss X Independent Variable (i.e. H vs. L) interaction. In other words, the experimental variable simply affects every S's performance by a constant increment but does not bring into play any other source of IDs.

In the multiplicative case $H = kL$. This is not detectable by the analysis of variance, which can yield a significant Ss X Independent Variable interaction even though the source of IDs is identical for both L and H. The correlation (Pearson r) between H and L, however, would be 1.00 (corrected for attenuation), the ratio of H/L would have zero reliability, and the correlation of H/L with any other variables (including H and L) would not be significantly greater than zero. A spurious ID phenotype, therefore, can be identified by these measures.

None of the above techniques, however, is adequate to detect the exponential case, where $H = L^k$. Here the analysis of variance (with a significant Ss X Independent Variable interaction) and the Pearson r (with r significantly less than 1.00), and the significant correlations between the ratio H/L and other variables can all give the appearance of a genuine source of IDs when in fact none exists. The rank order correlation (Spearman's rho) between L and H, however, will not differ significantly from 1.00 under this condition. Rho can therefore be used to rule out spurious IDs of the type $H = L^k$. When rho and Pearson r between L and H do not show an appreciable difference, this expondential relationship does not exist. In general, if Ss do not maintain the same rank order of performance on L and H, it is safe to assume that the two conditions involve different sources of IDs. These various tests have been applied to the present data to determine the authenticity of the ID phenotypes that are entered into the final factor analysis.

4. Establishing Reliability of Measurement. - A great deal of attention must be given to the problem of reliability of measurement in the study of IDs, for two main reasons. First, it is impossible to interpret a correlation matrix or the factor analysis thereof without knowing something about the reliabilities of the measurements that went into the correlations. If a particular test has low correlations with other tests in the battery and consequently has a small communality in the factor analysis, one wishes to know whether this is because what the test measures actually does have little in common with the other tests or simply because the reliability of the test is too low to allow it to correlate substantially with any other tests. In the second place, when tests differ widely in reliability, the tests of low reliability do not carry their weight in the identification of factors. Their loadings will necessarily be low on any factor and tests which may actually be less representative of the factor but which have higher reliability will outweigh the relatively unreliable test when it comes to the interpretation of the factor. Within limits one can correct factor loadings for attenuation if the test reliabilities are known. The best solution, of course, is to actually achieve substantial reliability of measurement on all tests so that correction for attenuation is either unnecessary for the unambiguous interpretation of the factors or at least involves little risk of contaminating the factor loadings with the error contained in the reliability estimate itself.

Since many of the correlations that enter into a factor analysis are between tests administered to Ss on different occasions, the reliability we are most concerned with is test-retest reliability. To rule out specific practice effects, the test-retest reliability was determined by administering equivalent forms of each test one week apart. If, for example, we wish to interpret the correlation between tests A and B when these have been administered on separate occasions, we need to know the reliability of both A and B based on test-retest on separate occasions. Split-half reliability or other estimates of internal consistency would be inappropriate for this purpose. In other words, our tests must have substantial stability in time if they are to be of use in the study of IDs, unless, of course, we are specifically interested in IDs in instability of performance from one occasion to another.

In the present study equivalent forms of each of the laboratory tests that were considered for factor analysis were administered to a separate group of Ss on two occasions one week apart. These groups are referred to hereafter as the Single-Test Groups. The purpose of these groups was to provide data for estimating the reliability of measurements derived from single tests and for "validating" the phenotypic IDs in interference that the tests were intended to measure. At the same time the data from the Single-Test Groups serve as a kind of "cross-validation" of the data obtained on the All-Tests Group on which the factor analysis was based. A large number of the intercorrelations among variables in the All-Tests Group are found also in the Single-Tests Groups, which can serve as a kind of check on the typicalness and generality of the results obtained in the factor analysis of the All-Tests correlations.

C. Factor Analysis of the Tests

All the tests in the battery were administered to one group of 50 Ss, hereafter referred to as the All-Tests Group. The final factor analysis, which is the culmination of this study, was based on the data of the All-Tests Group. Principal components, orthogonal and oblique simple structure rotations, and higher order factors were obtained from the zero-order correlations.

III. Related Research

The experimental literature involving the concept of interference is vast, while the literature on IDs in susceptibility to interference effects is strikingly meager. Though it would be far beyond the scope of the present paper to review the experimental literature on interference - a subject that could fill a whole book by itself - there are a number of lively issues in current research on this topic that are quire relevant to some of the methods and aims of the present study.

A. Distinction Between Long-Term and Short-Term Memory

First of all, research on interference can be roughly classified into two large categories: (a) interference in short term memory (STM) and (b) interference in long term memory (LTM). The present study takes account of this distinction by including tasks involving interference in both STM and LTM. At present the distinction between STM and LTM is rather arbitrary and does not rest firmly on any factual psychological difference between the two. Whether STM and LTM involve different psychological or neurophysiological processes or merely represent an arbitrary procedural distinction while the basic processes or mechanisms are the same for both is at present an unresolved theoretical issue. A majority of investigators in this field favor a two-process theory (e.g., Broadbent, 1958, 1963; Brown, 1964; Gerard, 1963; Wooldridge, 1963; Young, 1964), though strong arguments have been made for a uniprocess theory (e.g. Melton, 1963).

It is not the purpose of the present study to deal with this issue, but it is suggested that the study of IDs in LTM and STM might be a fruitful approach to the further investigation of the problem. While a high correlation between IDs in STM and LTM would not necessarily rule out a dual process theory (since two processes can be different though correlated), a lack of correlation between IDs in STM and LTM, assuming other factors are properly controlled, would be practically fatal to a uniprocess theory, unless, of course, one rejects the reasonable assumption that a common process should give rise to correlated IDs in measures that involve this process. So far it is the neurophysiological, rather than the purely behavioral, evidence which seems to weight the balance in favor of the dual process theory. Young's research on the octopus, for example, has shown that, at least in that animal, LTM can be literally extirpated from the brain (a separate lobe, in fact), leaving STM unimpaired (Young, 1964). Pathological conditions of memory also seem more consistent with the dual process view. Patients with Korsakoff's syndrome, for example, can perform within the normal range on a test of immediate memory span and yet half an hour later will have no recollection even of having taken the test. Such phenomena suggest some discontinuity in memory processes.

STM and LTM can be operationally distinguished for experimental purposes, however, without involving any theoretical issues. Melton (1963) has defined STM as memory over intervals up to 5 minutes, which seems an overly liberal criterion. Mackworth (1964) has made a further distinction, using the term Very Short-Term Memory (VSTM) as memory for material which has been presented long enough to be fully recognized and named, with delays between presentation and recall of each item of the order of 10 seconds or less. Perhaps the most comprehensive distinction between STM and LTM, and the one that completely corresponds to the distinctions made in the present study, is that given by Wickelgren (1965): "STM is being studied when the list is presented once at a rate of less than 2 sec per item and retention is assessed after less than 30 sec, and LTM is being studied when the list is presented several times over a period of time measured in minutes and retention is assessed after minutes, hours, or days."

By definition interference always results in forgetting or a performance decrement of some kind. One major theoretical issue concerns whether interference operates by the same or by different processes in STM and LTM. The literature provides no consistent answer to this question. Some investigators have held that interference must act in different ways in STM and LTM on the ground that degree of similarity of the interpolated stimuli and the to-be-recalled stimuli is a crucial variable in LTM and is relatively unimportant in STM (Brown, 1958, 1964). Wickelgren (1965), on the other hand, has shown a close parallel between LTM and STM with respect to forgetting as a function of similarity of the interpolated material. Degree of acoustic similarity, rather than visual or semantic similarity, proved a potent variable in STM. Thus, at least for the present, STM and LTM do not seem to be clearly distinguishable in terms of the effect of similarity on the amount of RI.

An even more unresolved issue concerns whether forgetting in STM is due to spontaneous decay of the memory trace or to active interference with the consolidation of the memory trace by other stimuli. Since it has so far been impossible to shut out extraneous, interfering stimuli (external or internal) during the interval between presentation and recall, trace decay has never been detected in pure form. If the S's activity during the interval between presentation and recall is not controlled by the experimenter, it is presumed that the stimulus trace undergoes consolidation, rehearsal, or recirculation through the STM store. The terms "consolidation", "rehearsal," and "recirculation" are all hypothetical constructs, and as far as this writer can make out there is no consistent functional distinction between these terms in the literature. All that they seem to mean is that the more control the experimenter has over the S's attention during the interval between presentation and recall, the less the S can recall. The intervening demand on attention is presumed to interfere with whatever process takes place under less attention-demanding conditions that result in greater recall. The writer refers to this hypothetical process as consolidation. The term "consolidation" is preferred to "rehearsal" since the latter implies a conscious effort on the S's part. If conscious rehearsal were necessary for the persistence of the trace it is hard to see how infants and lower animals would learn. Consolidation is thought of as an autonomous neural process. Rehearsal is thought of here as a learned, consciously adopted technique by which the S tries to protect the memory trace from interference. Improvement of the memory span with practice (e.g. Gates & Taylor, 1925) probably results in part from the acquisition of improved techniques for minimizing interference.

While it is not known whether the stimulus trace decays in time without interference and without rehearsal or consolidation (which presumably take time), it is very clear that intervening activity, even for extremely short durations, can result in large losses in recall. Conrad (1960), for example, presented Ss with a series of 8 digits and required them to speak the word "zero" before recalling the series. Though uttering "zero" took only a fraction of a second, it resulted in a very appreciable increase in the frequency with which Ss failed to recall the 8 item series entirely correctly, as compared with the control conditions, which did not require any intervening response between presentation and recall. The great susceptibility of immediate memory span to interference of this type makes it an ideal task for the measurement of IDs in interference proneness.

A purely logical analysis of the problem leads to a hypothesis of two basic types of interference. (The tenability of the hypothesis is, of course, an empirical matter.) If we assume that a stimulus trace can be strengthened (without repetition of the stimulus) or maintained throughout a period of time after the presentation of the stimulus, and if we use the term "consolidation" to describe whatever the process is that strengthens or maintains the stimulus trace, then we can say that, during the course of learning, a stimulus trace is in one of two states: it is either in the process of being consolidated, or it is at the asymptote of consolidation. As the state of the trace progresses from the beginning of consolidation to the asympotote of consolidation its resistance to certain kinds of interference decreases. Since the trace can be conceived of as existing in one of two different states, one can hypothesize that interference operates differently in the two states. Interpolated activity could interfere with the process of consolidation only while consolidation is taking place. The stimulus trace would be "erased" or weakened through such interference, and it would presumably be unrecoverable. After consolidation was complete, however, interpolated activity would have to exert its damaging effect through some other means than interference with consolidation.

The processes hypothesized to cause forgetting in LTM are response competition and extinction. Response competition comes into play when there is some similarity between the original learning (OL) and the interpolated learning (IL) and the IL acquires an associative strength approximately as great as that of OL. Under these conditions the two sets of responses, OL and IL, compete and interfere with one another in recall of either OL (which is known as reactive inhibition) or IL (known as proactive inhibition), resulting in errors of intrusion of items from the "wrong" list and in "blocks" or failures to respond within a specified time-limit. Response competition of this type, however, is not the whole story. In the RI paradigm there is an actual loss of the availability of items in OL after IL, shown by the fact that when the subject is required to recall both OL and IL items, there is a loss of OL items. *It thus appears that in the RI paradigm both OL and* IL do not remain side-by-side fully intact to compete with each other in recall purely on the basis of relative response strengths. Rather it appears that OL has also been weakened during the course of IL. This weakening of OL has been likened to Pavlovian extinction and attempts have been made, not with unambiguous success, to demonstrate spontaneous recovery of the OL responses over time. These two princip - competition resulting from response dominance and weakening of responses through extinction - are the essence of the interference theory of long-term forgetting. A detailed evaluation of the theory has been presented by Postman (1961).

Another way in which LTM differs from STM is in terms of retrieval of the stored memory. In STM there seems to be no process of retrieving the material to be recalled; indeed, if it is not recalled within a rather brief interval, it seems to be lost for good. In LTM, on the other hand, everyone has had the experience of not being able to recall immediately on demand some bit of information, like someone's name, that he is certain he knows. And usually, after a few moments, the information being sought will come to mind. So there may also be a factor of IDs in retrieval of information in LTM which would not enter into performance on STM tasks.

The main point of this discussion is that there seems to be some justification on the basis of current theoretical formulations for hypothesizing different interference processes in STM and LTM. One aim of ID research is to determine whether IDs in the hypothesized processes (interference with consolidation, response competition, and extinction or "unlearning") are at all independent.

B. Interference in Immediate Memory Span

Immediate memory span is highly subject to interference effects, and since much use is made of this fact in the present experiments, it might be worth while to discuss briefly the main intrinsic sources of interference in memory span tasks. The usual memory span test consists of presenting a series of items one at a time at a uniform rate. The subject (*S*) then recalls the series in the order of presentation. To do this completely successfully the *S* must have resisted to a large degree interference arising from (*a*) successive presentation, and from (*b*) reproduction of the items.

First, it is known that successive stimuli can "erase" previous ones (Averbach & Sperling, 1961). In fact, if the interval between the first and the second item is less than about 1/5 sec, the second item will completely erase the trace of the first (Gilbert, 1959). As the interval between items increases, more consolidation of the trace can occur and the trace becomes more resistant to interference from successive items. Even within the 1 sec interval between items in the usual memory span test some interference takes place. As will be shown in the present study, even as few as two items (digits) presented at a 1 sec rate will evince a slight but significant interference effect. The interference that is wrought by the presentation of each successive item increases through the series, because what must be rehearsed or consolidated in any 1 sec. interval is not just *the single item presented in that interval* but the whole series up to, and including, the item then being presented. If, before the presentation of each new item, the series were more or less completely consolidated up to that point, interference from successive presentation would be minimal and the *S* should be able to recall a much longer series. To take an extreme example, it would be as if the first seven digits in the series were the *S*'s telephone number; in such a case the *S* might well achieve a memory span of 14 or 15 digits.

Immediately after the series has been completely presented, it is known that its representation in the *S*'s memory is much more complete than would be suggested by the *S*'s performance when he tries to recall the whole series from beginning to end (Sperling, 1960). The loss in retention that occurs between the end of presentation and the time of recall is theoretically attributable to (*a*) simple decay of the memory trace between the time of presentation and the time of recall, (*b*) interference with consolidation due to successive presentation of items, and (*c*) reproductive interference with items later in the series by the act of recalling preceding items. Reproductive interference seems to be the most potent cause of forgetting in memory span tasks. Whatever other types of interference may be experimentally introduced into memory span tasks, it is unavoidable that all such tasks will be loaded with at least these two forms of interference, i.e. interference from successive presentation and from reproduction. Simultaneous presentation of the series and post-stimulus cuing (requiring recall of only a limited portion of the series) can, however, reduce these two types of interference effects. Though these techniques were not used in the present study, it is likely that their use will be required in further, more refined and analytical studies of IDs in memory span.

In addition to these interference effects, another factor may enter into IDs in memory span, *viz.*, the strength of immediate registration of the stimulus trace. If the stimulus leaves a trace that must be consolidated or which can be erased by interference, it is reasonable to ask if there are intersubject and intrasubject differences in the strength of the stimulus trace immediately upon the cessation of the physical stimulus. Differences in strength of registration could be one source of variance in memory span as well as rate of consolidation and resistance to interference with consolidation. Thus we see that there are hypothetically at least three sources of variance in memory span. The process might be likened to that of photography. The impression on the photographic plate can be strong or weak, depending on the sensitivity of the plate. The image is then developed (consolidated) slowly or rapidly, depending on the concentration of the developing solution. And during the course of development the image can come out sharply or dimly, depending on how much stray light (interference) strikes the film between the time the picture was taken and the time the film was completely developed. Since memory span is highly related to general intelligence, we might discern in these three processes the basic functional ingredients of intelligence. The same would hold true for learning, since every S-R contingency can be presumed to undergo consolidation and to be liable to interference effects. Rapid consolidation would mean that more could be learned per unit of time; it would remove the trace from interference more quickly. Similarly, resistance to interference would result in more being learned per unit of time. Furthermore, weak registration of initial input would mean that there was less trace to be consolidated and that less interference would be needed to erase the trace. Thus, the seemingly small individual differences one finds in immediate memory span can, when multiplied over a lifetime, make for large differences in "intelligence." It is known, for example, that *Ss* whose memory spans differ by only a single digit will have vocabularies that differ by several thousand words. (The correlation between the Vocabulary and Digit Span tests of the Wechsler Adult Intelligence Scale is .74. Corrected for attenuation the correlation is .93).

As a final word about the role of interference in immediate memory span we should now clear up a possible misconception. A person's memory span is not to be viewed as an independent capacity of the person which can be interfered with only by external means. The span itself is seen as being intrinsically a function of rate of stimulus trace consolidation process. In other words, a person's memory span is not to be thought of as *x* number of items which suffer interference only when this number is exceeded. The writer has not found any evidence of any kind of "magical number seven" or any other "magical" number for any individual or group of individuals which is marked by any clear-cut discontinuity in probability of recall (see Miller, 1956). As was originally pointed out by Guilford and Dallenbach (1925), probability of correct recall is represented by a perfectly smooth ogive which connects the length of series at which recall is always perfect with the length which the S never recalls correctly. In fact, the probability of correct recall of a two-item series is less than the probability of correct recall of a single item. Each item added to the series after the first decreases the probability of correct recall. What we call the S's memory span is an arbitrary point on the ogival function mentioned above; the point usually taken is that at which recall is correct 50 percent of the time. Further analysis reveals that the probability of recall of single items in the series is a function of their serial position, so that it is the most "difficult" item in the series which determines the S's memory span by this arbitrary cut at 50 percent. By this definition the S's memory span is the length of series in which the most "difficult" item in the series has a recall probability of .50. This method of measuring memory span actually uses only a fraction of the information in the S's performance and thus seems unnecessarily crude for the study of IDs. Other methods of scoring memory span performance have therefore been used in the present study. The success of these methods is evidenced by the very high reliability of the scores, even within a relatively homogeneous population with respect to age and ability, and by the highly significant correlations of the scores with external criteria.

In summary, the main point is that there seems to be no evidence that the effect of interference on memory span is at all analogous to the straw that breaks the camel's back. The influence of interference is a continuous, though not a linear, function of the number of items presented. The information theory concept of "channel capacity" seems to this writer to be a too static and passive mechanism to explain the facts of memory span. The concept of interference, on the other hand, seems almost impossible to avoid in accounting for some of the facts of memory span. One example is the fact, originally pointed out by Gates (1916), that when the number of items in the series presented to the S exceeds the S's memory span (defined as perfect recall 50 percent of the time), even by as much as one item, the number of items the S is able to recall is less than his span. Each additional item interferes with and diminishes the traces of the preceding items. This effect is made use of for measuring IDs in interference in the present study.

C. Interference and Individual Differences

1. Age. - In the literature on IDs in learning, interference effects have been referred to most frequently in relation to aging. This is especially true with respect to STM. Cameron (1943) found that senile persons suffer greater retroactive inhibition of STM than do young adults. And in a review of the literature on the effects of aging on human skills Welford (1958) concluded: "Increased liability to interference with short-term retention by other activity would seem capable of accounting, directly or indirectly, for most if not all of the cases where true learning or retention deficiency can be shown, even those where other explanations seem at first sight more obvious" (p. 256).

The fact that memory span increases steadily with age from early childhood up to about 18 years of age and then gradually begins to decline at a rate that parallels the decline of a number of physiological measures, such as vital capacity, suggests at least two things. First, it suggests that in our work with memory span we are dealing with a means of assessing some quite basic aspect of the functioning of the nervous system, in the sense that the decline of memory span is closely related to decline of a number of physiological measures (Welford, 1958). Secondly, it suggests we are dealing with a type of performance that is not predominantly influenced by learning and practice. It would be almost as difficult to account for the decline of memory span in experiential and nonphysiological terms as to account for the closely parallel decline in vital capacity in such terms. It therefore seems reasonable to assume that in measuring IDs in memory span and the effects of interference on memory span we are getting at quite fundamental aspects of brain functioning.

That the effects of aging on immediate memory may not be an entirely simple matter, however, it indicated by an experiment by McGhie, Chapman, and Lawson (1965), who compared three age groups (under 40, 41-60, over 60) on a STM task under conditions both of visual and of auditory presentation. Retention of both visual and auditory material declined with age, but visually presented material showed a greater decline than aurally presented material; the diffference between the modalities was most pronounced in the over-60 age group. The technique used by Lawson, incidentally, was such as to minimize reproductive interference; practically all the interference in this experiment was due to RI from successive presentation of digit pairs. One pair of digits in each series was repeated; the S's task was to write down only the repeated pair of digits after the entire series (10 pairs) had been presented. This kind of finding, as well as others to be mentioned later, made it advisable in the present study to investigate the effects of auditory versus visual modes of presentation. It can be said in advance that while auditory and visual presentation showed some highly significant differences as well as different interactions with experimental variables, sensory modality showed not the least interaction with Ss in this study. It could be that age is the only subject variable that interacts with sensory modality of stimulus presentation in STM. One would therefore not expect much interaction to appear when all the Ss are within a very narrow age range, as in the present study. At least there is some comfort in the fact that among the normal college population sensory modality does not seem to be an important source of IDs in STM.

The effects of aging on LTM are perhaps even more complicated. But the evidence is much less adequate than in the case of STM. It is quite clear that learning *rate* for most forms of laboratory learning decreases with age after about the age of 18 or 20 (for comprehensive reviews of this evidence see McGeoch & Irion, 1952, and Welford, 1958). Measures of retention, however, even after retroactive interference has been produced experimentally by an interpolated learning task, have not shown a significant relationship to age. An experiment by Gladis and Braun (1958) used Ss over an age range from 20 to 72 to investigate the effects of age on RI in rote learning. The three age groups (20-29, 40-49, 60-72) differed significantly in learning rate in both OL and IL. But they did not differ significantly in retroactive interference (RI) as measured by both recall and relearning, adjusted for differences in vocabulary and level of learning ability as measured in OL and IL. A likely hypothesis is that aging slows down the rate of consolidation, which would have the effect of slowing down learning and the S would be required to take more trials to consolidate the material. Once the material is consolidated or is in the LTM store, the scanty evidence suggests that age is not an important source of IDs in retention. It would seem that speed of retrieval should also decline with age, but to the best of the writer's knowledge there has been no investigation of this specific aspect of recall in relation to age. Another way that age could have its effect on learning would be through the strength of initial registration of the input. But it remains for future research to sort out the various loci of the effects of aging on learning and retention in both STM and LTM.

2. Interference in Slow and Fast Learning. - If susceptibility to interference effects is an important factor in speed of learning, we should expect fast and slow learners to differ in their susceptibility to interference, even if interference is only one of a number of factors determining speed of learning. Stroud and Carter (1961) investigated this matter by selecting Fast and Slow learners (top and bottom 15 percent) on two paired-associate learning tasks. The groups were then compared on speed of learning short (12-item)and long (24-item)paired-associate lists. The long list is presumed to create greater intralist interference than the short list. The Fast learners took on the average 9.16 trials to learn the short list and 36.38 trials for the long list. The corresponding figures for the Slow learners were 36.00 and 114.88. The absolute increase in learning time on the long list for the Fast learners is clearly much less than for the Slow learners, but the relative effect of list length on the two groups was about the same. This is what would be expected if the length-difficulty relationship is linear within this range of list lengths. Beyond the first few items in the list the amount of intralist interference exerting an influence on successive items is probably already at an asmptotic value, so that an increase in list length from 12 to 24 items could not be very effective in decreasing the rate of acquisition of single items. Another feature of this experiment suggests that the retention of already-learned items is more readily interfered with in Slow than in Fast learners. In both the 12-item and 24-item lists there were interspersed three paired-associates that the Ss had acquired to a common criterion in a previously learned list. The Fast learners were better able to withstand the interference from the new, unlearned items and held on to the previously learned items significantly better than the Slow learners. It did not matter whether the learned items were included in the 12-item or 24-item list; the retention loss was the same, again probably because for most Ss intralist interference reached an asymptote within 12 items. Consistent with other findings (e.g. Stroud & Schoer, 1959, and Underwood, 1954), the Slow and Fast learners did not differ significantly in a recall test administered 48 hours after the original learning. If rate of learning is uncorrelated with degree of long-term retention, as the best available evidence indicates, it suggests that different processes underlie IDs in STM and LTM.

The effect of distribution of practice on verbal learning probably interacts with IDs in susceptibility to interference in determining speed of learning. Underwood's extensive research on the conditions under which distribution of practice facilitates verbal rote learning has revealed that the critical variable determining the amount of facilitation gained by distributed as opposed to massed learning trials is the degree of intralist similarity (Underwood, 1953). Greater similarity means greater interitem interference. The longer intertrial interval of distributed practice presumably allows the traces laid down in each trial to undergo a greater degree of consolidation during the intertrial interval than would occur under massed trials. (By the same token we should expect pacing rate to interact with intralist similarity even more strongly than distribution of practice.)

An experiment by Madsen (1963) is the first demonstration with human *Ss* of an interaction between a subject variable, in this case IQ, and distribution of practice. The results seem consistent with the idea that the intertrial interval allows time for consolidation to take place. If the interval is long, fast and slow learners (consolidators) should both attain a high degree of consolidation of the traces after each trial and the difference in their learning rates should be relatively small. With a short intertrial interval, on the other hand, fast learners would be expected to consolidate more of the traces than slow leraners, and the difference in their learning rates should be correspondingly great. This is essentially what Madsen found. Three groups of children of high, average, and low IQ learned 10-item paired-associate lists under massed (5 sec. between trials) and distributed (60 sec. between trials) practice. (The intertrial interval was filled by color-naming.) It was found that the lower the IQ, the greater the benefit from distributed practice. The lowest IQ group required almost twice as many trials to attain criterion under massed as under distributed *practice, while the highest IQ group did slightly,* though not significantly, better under massed than under distributed practice. Thus there was a reversal of the effects of distribution of practice as a function of IQ level. Our understanding of the fundamental nature of intelligence will probably be advanced more in the future by experiments of this sort than by attempts to further refine psychometric tests of the paper-and-pencil variety.

IV. Procedures

A. *Subjects*

Subjects were volunteers from introductory courses in Educational Psychology and General Psychology in the University of California, Berkeley. Almost every undergraduate major offered in the University is represented among this sample. A total of 530 *Ss* served in the project; this represents a total of approximately 1000 subject-hours of individual laboratory testing.

The mean age of Ss in various experiments ranged from 19.4 to 24.5 years, with SDs ranging from 1.3 to 6.6 years. The overall mean was 22.7 years, SD = 3.6 years. The All-Tests Group (N = 50) had the lowest mean age and the smallest SD: mean = 19.4 years, SD = 1.3 years.

The percentage of males in the various experiments ranged from 14 to 53, with a mean of 25 percent. The All-Tests Group had 30 percent males. The reason for the preponderance of females, of course, is that most of the Ss were drawn from undergraduate courses in Education which are required for teaching credentials. It should be noted, however, that there is no undergraduate major in Education at the University of California and all of the Ss drawn from Education courses represent a large variety of undergraduate majors, as is also true of the Ss drawn from the introductory course in General Psychology.

A question might be raised concerning the suitability of university students as Ss in experiments on individual differences in learning abilities. The use of such Ss has both advantages and disadvantages. The most obvious disadvantage is the fact that, judging from the criteria for admission to the University and from the intelligence test scores on Ss in the present experiment, we are sampling Ss from the upper five to ten percent of the general population in intellectual ability. Consequently, the range of individual differences on our laboratory learning tasks will be restricted in proportion to the degree to which performance on these tasks is correlated with general intelligence. This should make it more difficult to demonstrate highly reliable individual differences in learning and memory among college students. If highly reliable measurements of individual differences can be achieved under these conditions of restricted range of talent, however, it seems safe to assume an even greater spread of scores and still higher reliability could be obtained in the general population. The difficulties of obtaining highly reliable measurements within a restricted range of talent were very satisfactorily overcome in the present experiments, though not without a good deal of preliminary trial-and-error.

There are other advantages of using college students in the early stages of research on individual differences in learning aside from the most obvious ones such as their accessibility, their relative quickness in comprehending and following instructions, and so forth. Perhaps the main advantage with respect to experiments on cognitive abilities is the fact that in using college students we are dealing, not only with a superior segment of the population, but with Ss all of whom are comparatively free of abnormalities or handicaps in the cognitive sphere. The entire educational process, from kindergarten on, and especially the criteria for university selection and for subsequent academic success, acts as a very thorough form of screening out intellectual defects. Organic brain conditions, aphasias, dyslexias, "cultural deprivation," psychotic thought disorders, or other deviations from normal brain functioning are undoubtedly much rarer in university students than in the general population. These types of defects, while of great psychological interest in themselves, would only add "noise" to any attempt to delineate the dimensionality of individual differences as they exist in wholly intact, normally functioning individuals. At a later stage of investigation, when one wishes to test some model or hypothesis about what goes on in the "black box" in certain types of learning, it may be of value to study abnormal subjects, such as persons with specific, well-defined kinds of brain damage. The model would have to be revised, for example, if it were found to be inconsistent with any type of abnormality that can occur in the processes which the model is intended to explain.

One other problem of using a restricted sample of the population in a study such as the present one concerns the nature and generality of the "dimensions" derived from factor analysis. For our present purposes this is not a crucial issue; it will be considered, however, in the discussion of the results of the factor anlaysis.

B. Apparatus

The principal apparatus used throughout all the experiments was the "multiple S-R apparatus for human learning," which has been described in detail elsewhere (Jensen, Collins, & Vreeland, 1962). It consists essentially of an in-line display unit for the presentation of the stimuli and a response panel with an array of pushbuttons on which the S makes his responses. The timing and sequence of the presentation of stimuli were all programmed on punched tape. Thus, each experiment was run completely automatically after the S has received the preliminary instructions. This insured great uniformity of the procedure for all Ss.

The S worked alone in a "sound protected" room which was barren except for the simple apparatus essential for the experiment. The tape mechanism and control unit were in an adjoining room from which the experimenter could observe the S through a one-way mirror.

The apparatus was used in three slightly different ways for (a) the trial-and-error learning tasks which proved unsatisfactory and were abandoned, (b) the memory span experiments, and (c) the serial learning experiments. The specific procedures of each of these types of experiments are described in the following section.

A word is in order about the reliability of the apparatus. In the two years of running subjects -- well over 1000 hours of testing, since many subjects were run in other experiments not connected with this project -- there were only five or six apparatus failures, and nearly all of these occurred in the first three or four months of experimentation. After a new timing device (mechanical rather than electronic) was installed, there were no further difficulties whatever. This is a remarkable record for a rather complicated electronic apparatus which was required to present a stimulus once every second for a total of over 1000 hours.

C. Unsatisfactory and Abandoned Techniques

A number of the procedures described in the original proposal of this project, as well as variations of these procedures, were attempted and were found to be unsatisfactory for the purposes of the present study. Somewhat simplified versions of these unsuccessful techniques had been used successfully in previous experiments on school children and on mentally retarded adults without any suggestion of the kinds of difficulties that were to arise when the techniques were applied to university students.

The basic problem was the difficulty in achieving reasonably high test-retest reliability. For the purposes of factor analysis, low reliability of the measures must be compensated for by using a very large number of *S*s. Even then, tests of low reliability can contribute only a small amount to the common factor variance, and this can make for ambiguity and difficulty in the interpretation of factors. Also, a large number of *S*s is practically prohibitive when every *S* has to take a number of very time-consuming, individually administered tests. Therefore, unless a test had reasonably high reliability it could not be considered suitable for the present study. Furthermore, if tests are being developed for the study of individual differences, it would seem highly desirable that such tests have sufficiently high reliability to permit their use for individual diagnosis.

Since it might be of some value to other investigators in this area, if only to alert them to some of the possible difficulties, brief descriptions of the initial mistakes and failures in the present study and the reasons for them are given below.

All of these first attempts were conducted as pilot studies, some of them involving several changes in instructions and other aspects of procedure. In some cases it was apparent that a particular procedure was going to prove unsatisfactory after testing only five or ten *S*s. In other cases the inadequacy of the technique was not apparent until twenty or thirty *S*s had been tested and the results had been statistically analysed. Only those procedures which were piloted to the point of finally seeming worthy of trying out on groups of at least 30 *S*s without making any further modifications in the procedure will be described here in any detail.

All of these pilot experiments made use of the Multiple S-R Learning Apparatus. In every case the *S*'s task consisted of learning, by trial-and-error, the associations between each of a number of stimuli presented successively on the display screen and each of a corresponding number of pushbuttons on the *S*'s response panel. The "reinforcement" for a correct response always consisted of a green light, directly below the stimulus display screen, going "on". A record of the essential features of the *S*'s performance was automatically tabulated on the bank of counters on the control unit, which was monitored by a research assistant. As in all these experiments, the sequence and pacing of the stimulus presentation was programmed and automatically controlled.

1. *Retroactive Inhibition of Selective Trial-and-Error S-R Learning.* - At its conception this experiment seemed a beautifully simple and straightforward technique for measuring IDs in susceptibility to retroactive interference. Indeed, in an earlier study not concerned with the reliability of IDs, it had been found that this technique was capable of producing on the average a large amount of retroactive interference. But after several modifications of the instructions and scoring methods, using 70 *Ss* in all, the test was abandoned as practically hopeless for obtaining reliable measures of IDs, at least in the present college population. Test-retest reliabilities of the original learning (OL), interpolated learning (IL), and relearning (RL) parts of the test, based on 30 *Ss* who were put through the final version of the procedure, were not significantly greater than zero, ranging between .15 and .25.

The task consisted of three parts: OL, IL, and RL. In OL the S learned ten S-R connections (stimuli were digits from 0 to 9, responses were ten pushbuttons in a horizontal row labeled A to J). The IL task consisted of one of two conditions: a High interference condition or a Low interference condition. (Every *S* had one or the other of these conditions, but never both, since the aim at this point was only to determine the test-retest reliability of each condition separately. *Ss* were retested on a formally equivalent form of the task within a one-week interval.) The High interference IL condition consisted of learning a new set of S-R connections in which the stimulus and response items were identical to those in OL but were simply re-paired for the IL. This condition should (and does) produce maximum retroactive interference with the retention of the OL task. The Low interference condition of IL consisted of learning new S-R connections, this time with the stimuli consisting of consonant letters of the alphabet from K to Z. This condition produced relatively less interference than the re-pairing of the number-letter pairs used in the High interference condition. It is somewhat equivalent to the "control" or "rest" condition in the typical paradigm of experiments on retroactive inhibition.

The relearning (RL) task, which was intended to reflect the amount of retroactive interference (RI) due to the IL, was, of course, always identical with the OL task. If the measures of RL following both High and Low IL showed high test-retest reliability, it should then be possible to give both interference conditions to the same *Ss* and determine a percentage of RI for each *S*, based on the *S*'s RL scores after High and after Low interference from the IL. Unfortunately, as was pointed out, none of the measures obtained from these tasks was reliable, although, interestingly enough, the average results of the group were quite in line with expectations, showing an appreciable amoung of RI (about 30 percent).

After several revisions the final set of instructions to the *Ss* on this task were as follows:

a. On this screen you will be shown some digits. They are presented one at a time, at a constant rate of one every three seconds. The order of presentation is entirely random.

b. Down here are ten pushbuttons. Each button corresponds to one of the digits.

c. Your task is to learn, by trial-and-error, which button goes with each digit. When you press the "correct" button, this green light goes on. In this session you will learn one task, then you will learn a different task, and then you will go back to the first task to see how well you remember it.
d. Start with this strategy: Use button A until you discover which digit it goes with; then go on to Button B, and so on.
e. Your score is based on how quickly you learn the task as well as the percentage of the time you can make the green light go on.

Why did this technique fail to yield reliable measures? The reasons seem to fall into two categories, which we can call *criterion* problems and *strategy* problems.

The criterion problem was that of deciding the optimal criterion of learning for OL and IL for the purpose of measuring retroactive interference in RL. Since the order of presentation of stimuli was entirely random within blocks of ten, it was possible to set up as the criterion of learning the achievement of ten consecutive correct responses within a block of ten different stimulus presentations. This turned out to be a much too rigorous criterion; many *S*s never attained it in a reasonable length of time (say, 20 minutes). There was also the problem that some *S*s highly overlearned certain S-R connections on which they were consistently reinforced on trial after trial and more or less consistently failed to perform at better than a chance level on other S-R connections. No matter what criterion was adopted as the stopping point for OL or IL, it seemed impossible to regard all the *S*s as having attained an equal degree of learning. There seemed to be too many ways of arriving at the same objectively defined criterion. A rather novel criterion which was tried, but also without much success, was to stop when the *S* reached the point in learning when, after he had been reinforced at least once in each S-R connection, the total number of *correct* responses made by the *S* was equal to exactly 50 percent of all of his responses up to that point. The *S*'s score was the total number of stimulus presentations needed to reach this point. All *S*s were able to reach this criterion in a reasonable time. The main disadvantage of this criterion was that *S*s would attain it by quite different means. Two *S*s, for example, could end up with the same score, but there was no way of knowing if their scores represented an equivalent degree of learning. One *S* might have attained the criterion by completely mastering only four or five of the S-R connections, responding at a near chance level to all the others, while another *S* might attain a more or less uniform, better than chance level of responding to all ten of the items without ever achieving mastery on any one of them. The actual test-retest reliability of these learning scores, being practically nil, suggested that *S*s were not being consistent in the particular method by which they attained the criterion.

This brings us to the problem of the variability in *S*s' *strategies* on the S-R task. While the topic of strategies of attacking learning tasks may be of great psychological interest and importance in itself, strategy types of behavior are simply "noise" as far as the present experiment is concerned. Instructions to the *S*s were able to cut down the range of strategies to some extent, but never enough. Unlike most of the children and the mentally retarded adults we have tested in this fashion, college students could not resist "experimenting" with various strategies of attacking this S-R task, and many *Ss* not only changed their strategies in going from OL to IL or RL, but would also change their approach even in the midst of any one of these tasks. One approach, for example, was to concentrate on just two or three S-R connections at a time until they were overlearned before trying to obtain more information by testing other pushbuttons. *Ss* using this strategy would often avoid responding to stimuli for which they didn't know the "correct" response. Then there was the opposite extreme, a "risk-taking" strategy in which the *S* would seldom lose an opportunity to get information by testing a button every time a stimulus was presented. Many *Ss*, furthermore, overtly verbalized the S-R pairs, e.g. "Nine is C, six is F," etc. It would be interesting to know if on repeated testing *Ss* would finally settle on some particular strategy, how long it would take *Ss* to arrive at this point, if, indeed, it is ever arrived at, whether there are individual differences in the strategies finally arrived at, and whether the various strategies differ in effectiveness, both in terms of speed of learning and in terms of resistance to retroactive interference. These kinds of questions must be put aside temporarily, since they are not the major concern of the present investigation. The aims of the present study, in fact, could be best accomplished by reducing strategy behavior to the absolute minimum. Variance due to this source could be reduced, of course, by devising tasks on which nearly all *Ss* had the same mode of attack or at least in which *Ss* consistently maintained the same strategy throughout, even though differing from one another in the particular strategies they adopt.

Since intra-task changes in strategy seemed to occur less frequently than inter-task changes, an intra-task measure of susceptibility to interference in S-R trial-and-error learning was devised.

2. *Intra-task Interference in Selective Trial-and-Error S-R Lea[illegible]ng*. - In this experiment the *S* first learned six S-R connections to a criterion of ten correct out of twelve successive responses. At this point six more S-R units of a similar type were added to the original six and the *S* then had to learn all twelve connections. The measure of particular interest was the amount of decrement in performance on the first six S-R units that resulted from the interference of adding the second set of six S-R units.

The Multiple S-R Apparatus was used. The stimuli of both the first and second sets of stimuli were letters of the alphabet. The response panel exposed a circular array of twelve pushbuttons numbered clockwise from 1 to 12. The six relevant buttons for the first set of S-R units were unsystematically distributed among the twelve available pushbuttons. It was up to the *S* to discover which buttons would receive reinforcement in the first part of the experiment. The order of the stimulus presentation was random within sets of twelve. Other details of the procedure are spelled out in the instructions to the *Ss*, as follows:

a. On this screen you will see six letters of the alphabet, presented one at a time, every three seconds in a random order.
b. Each letter is connected to one and only one of these twelve buttons, ie., each letter is connected to a different button, so that a letter and a button forms a pair, which will always remain the same.
c. Notice that there are twelve buttons and only six letters. So six of the buttons *are* connected with the letters and six are *not*.
d. Your task is to learn which button goes with each letter.
e. Press *only one* button each time a letter appears on the screen. Do *not* press another button until a different letter appears, because after one button is depressed the panel is dead until another letter appears.
f. When you press the "right" button, this green light will go on.
g. It is very important that you follow this strategy in order to learn the connections: Start with button Number 1 and press it for each letter that appears on the screen until you discover which letter it goes with *or* until you discover that it is not connected with any of the letters. Then go on to button Number 2 and do the same. Then to button Number 3, and so on, through button Number 12. (Experimenter demonstrates). Of course, if a letter appears whose button you have already discovered, push that button. Your score will be the number of times you make the green light go on.
h. When you have learned which buttons go with each of the six letters, a bong will ring and there will be a short pause. Then the bong will ring again and letters will appear on the screen as before. But this time six new letters will be added, making a total of twelve letters. Nothing else will be changed: the original six connections you learned will still be good. Only now each and every button will have a corresponding letter, there being twelve letters and twelve buttons.
i. Your task will then be to learn the buttons corresponding to the six new letters, while at the same time retaining the six buttons you have already learned. The machine will stop when you have gotten a certain number right.
j. Any questions? When you hear the bong, be ready to begin.

These instructions may seem excessively detailed, but it was found that unless every point was made explicit college students would tend to fill in the missing information with their own hypotheses concerning the nature and conduct of the task, and this would greatly increase inter- and intra-subject variability in ways that are not of interest to the present investigation. In fact, some 20 *S*s were used to work out the above set of instructions. Less intelligent *S*s seem to need less explicit instruction. With the college students we at times noticed *S*s going about the task in completely strange ways that only defeated our purpose. A couple of *S*s, for example, got the strange notion that once a button was reinforced by the green light going on, they were not to push that button again if they could possibly avoid doing so. Again, strategy effects were prominent among the majority of *S*s. Some *S*s tested the buttons one after another in a highly systematic fashion; others pressed one button after another in what appeared to be a haphazard fashion.

The "undesirable" forms of behavior elicited by the task were reflectd in the results. A group of 30 *S*s was administered equivalent forms of the task one week apart, using the final version of the instructions presented above. The *S*'s "interference" score was the decrement in performance on the first six S-R units after the addition of the second six. Specifically, it was the number of misses or incorrect responses made on all six S-R units within the first twelve cycles (i.e., 72 opportunities for getting the original six units) after the addition of the second set of six S-R units. In the group of 30 *S*s this decrement score ranged from 1 to 42, with a mean of 14. The test-retest (or equivalent forms) reliability was .27, which is not significantly greater than zero for *N* = 30. Thus, this test had to be judged as unsuitable for the measurement of IDs. The reasons for its failure are much the same as those described in connection with the previous task involving retroactive interference. The present part-whole learning task did not overcome these difficulties. Indeed, the practically inescapable conclusion was that the pushbutton mode of responding was an irresistible invitation to college students to test out various strategies and to entertain various self-generated hypotheses about possible systematic solutions to the problem. One very bright *S*, a mathematics major, for example, showed almost no learning at all during the session since he approached the whole task as a problem to be solved, as a pattern or code to be "cracked", rather than accepting at face value the experimenter's description of the task as simple trial-and-error learning. He was so dismayed by his lack of success after 40 minutes that he refused to come back for the retest a week later. Interestingly enough, a very similar S-R task has been learned within 15 to 30 minutes by school children in the IQ range of 50 to 75!

3. *Interference Due to S-R Similarity*. - One more pushbutton task was tried, this time with a short preliminary practice session which it was hoped would convince *S*s that nothing more than simple trial-and-error learning was involved and which might also serve to get all *S*s using a uniform strategy. The preliminary practice task was closely observed by the experimenter, who would stop the *S* and amplify the instructions if the *S* showed any signs of misunderstanding the nature and requirements of the task. The practice task consisted of exposing only four pushbuttons which were paired with four stimuli (colored forms) that were used only in the practice session. The *S* was shown how to start with button Number 1 until it had been reinforced, then to go on button Number 2, and so on, always trying to go back to buttons that had been previously reinforced whenever their corresponding stimuli appeared. This worked fine and all of the ten *S*s who were tried out on this test learned quickly and complied with the rules without much need for correction. The trouble began when the actual test was presented, and that is why the whole thing was abandoned after only 10 *S*s had been run. The task consisted of having *S*s learn two sets of 10 S-R connections during the 50 minute session. In one test the stimuli were ten letters of the alphabet and the response buttons were labeled with

numerals from 0 to 9; this was the "low interference" condition. The "high interference" condition was achieved by pairing the stimulus letters (a different set than those used in the first set) with pushbuttons labeled with the same ten letters, though, of course, the "correct" pushbutton for a given stimulus letter was never labeled with that letter. The high interference condition appeared to be more difficult than the low interference condition. Since *S*s' behavior on this test was hardly different from that on the other pushbutton tasks, it was not deemed worth while to continue, and the number of cases was too small to warrant statistical analysis. Simple inspection of the *S*s' scores suggested that they were not likely to be any more reliable than those from previous tests. It was decided, therefore, to throw out the pushbutton technique altogether.

This outcome should not be construed to imply that multiple S-R trial-and-error learning, using the pushbutton mode of response, is hopeless with all *S*s in all types of experiments. Because of the simplicity of the response mode it has been found almost ideal for work with children and retardates and *S*s with special speech problems or other performance handicaps. The technique has been used successfully with such *S*s (e.g., Jensen, 1963). Unfortunately, it does not yield individually reliable scores in the college population and any basic learning processes of the type we are concerned with in the present investigation seem to be almost completely swamped by more complex forms of cognitive behavior usually involving highly variable problem-solving strategies.

The next section describes the alternative tests that were devised and proved successful after a little piloting to perfect the instructions and procedures.

D. Interference in Short-term Memory Tasks

Five different short-term memory tasks were used in the present study. One of these tasks employed both visual and auditory modes of stimulus presentation. Before describing each task in detail, the features common to all the tasks will be described.

The *S* sat at a table on which were the stimulus display screen and the response panel. The display screen was at *S*'s eye level at a distance of approximately three feet. The figures that appeared on the screen (digits and geometric forms) were about two inches in height. The surface of the response panel, directly in front of the *S*, was inclined at a 30-degree angle from the horizongal plane. In the center of the panel was a clip-board, which held the *S*'s "answer" sheets. About three inches on either side of the clip-board was a pushbutton; the one on the right was labeled with a "+" sign, the one on the left was labeled with a "-" sign. The *S* was always instructed to keep his fingers poised over these pushbuttons during the presentation of stimuli; this was intended to prevent the *S* from writing during the stimulus presentation.

Ss were provided with specially prepared "answer" sheets, each 8 1/2" x 11" sheet containing eight horizontal rows of "boxes" for writing down each series of stimuli after it was presented. The boxes were 1/2 in. square and in each row were directly adjacent to one another. The number of "boxes" in each row was always the same for a given task, but ranged from seven to twelve for the various tasks.

All tasks were designed to take 35 to 40 minutes, leaving time for instructions, a short practice test, the administration of the Stroop Test, and filling out a brief personal data sheet, all within the one hour scheduled for each test session.

The two "equivalent forms" of each test were identical in every respect except the sequence of stimulus presentations.

In addition to the verbal instructions from the experimenter, each test was preceded by a brief practice test, taking about 10 minutes, identical in form to the actual test. The *S* was carefully observed during the practice test to insure that he was following the instructions properly. The practice series could be interrupted at any point whenever additional instructions seemed necessary. *Ss* were not allowed to begin the main test until it was clear to the experimenter that they were thoroughly familiar and at ease with the testing procedure.

A special signal, a green disc, used for no other purpose in the experiment, appeared on the display screen for ten seconds at the end of every page of the answer sheets; this served to remind the *S* to remove the top answer sheet and to be ready for the next series of tests. The answer sheets were numbered conspicuously in the upper right corner and were attached to the clip-board in the proper order.

During the main test the experimenter was never in the room with the *S*, but was able to observe the *S* through a one-way window from an adjoining room. External forms of distraction were thus kept to a minimum.

Labeling Tests. - Since the various tests about to be described will have to be referred to repeatedly throughout this report, it will be convenient to adopt some consistent set of abbreviated labels for the tests. Therefore all tests will be designated by the initial letters of their full descriptive labels plus a set of numbers indicating the lengths of the stimulus series used in each test.

1. *Immediate Digit Span (IDS, 2-9) and Delayed Digit Span (DDS, 2-9)*. - In the first test (IDS, 2-9), the *S* would see a series of digits presented singly at a one-second rate and would write down the series on his answer sheet immediately after the presentation. In the second test (DDS, 2-9), the *S*'s recall was delayed by 10 seconds. The delay interval was filled by the presentation, at a one-second rate, of red plus (+) and minus (-) signs in a random order. There were eight signs in all, and the *S* was required to push the corresponding pushbuttons labeled "+" and "-" as each sign appeared on the screen. These responses were tabulated on counters in the control unit, so it was possible to tell if

*S*s were obeying instructions. Almost without exception *S*s were conscientious in attending to the "+" and "-" signs and in pushing the buttons in accord with the instructions. ("+" was always under the right hand, "-" under the left hand.) In all cases 13 seconds were allowed for the *S* to write the digit series.

The IDS and DDS series were randomly interspersed within sets of eight; there were ten such sets in all. Each answer sheet provided space for eight series.

The digit series varied in length from two to nine digits. Each length of series was replicated five times throughout the entire test. To summarize:

2 Conditions (Immediate Recall and Delayed Recall)
8 Series Lengths (2, 3, 4, 5, 6, 7, 8, 9)
5 Replications

The exact procedure was as follows. The task was explained to the *S* with the aid of a wall chart which showed the two types of conditions and the design of the procedure. Each series was preceded by a bong as a signal to pay attention to the display screen. One second after the bong the digit series began. In the IDS condition the bong sounded immediately after the last digit in the series went off the screen; this terminal bong was the signal to write down the series in the appropriate space on the answer sheet. In the DDS condition the digit series was followed by a one-second "blank" interval, immediately followed by the red "+" and "-" series (eight in all), followed by the bong. The bong always filled a one-second interval. Thirteen seconds elapsed before the bong again signaled the beginning of the next series. An additional ten seconds, during which a green disc would appear on the screen, was always given between the pages of the answer sheets, to allow the *S* to remove the top sheet and be ready for the next digit series. To summarize the sequence.

Immediate Digit Span:

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
digits (2 to 9)	2-9 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

Delayed Digit Span:

bong (ready signal)	1 sec.
blank	1 sec.
digits (2 to 9)	2-9 sec.
blank	1 sec.
+ and -	8 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

It should be pointed out that in this test, as in all the other memory span tests used in this study, the two experimental conditions randomly interspersed within sets of eight throughout the test session. This procedure had two main advantages: (a) practice effects were more or less evenly distributed over the two conditions, and (b) the random interspersion of the two conditions would not permit *S*s to know during the presentation which condition they were getting, so they could not adopt different attitudinal or attentional sets while viewing and "encoding" the series for immediate and delayed recall.

The test proper was preceded by a practice test of eight series (four of each condition); the digit series were never longer than five in the practice tests.

The "equivalent" form of this test, as with all the other memory span tests, was identical in format. The only differences were in the digit series themselves.

The digit series were obtained in the following manner. The IBM 7090 computer made up and printed out several hundred sets of 9-digit numbers using each digit only once. (Zero was never used.) The sequence was always random with the one exception that no two digits were ever in the normal numerical order in the forward direction; thus, sequences such as 3-4 or 7-8 did not occur.

1A. *Auditory Form of the Immediate and Delayed Memory Span Tests*. - Since there was a question of whether the results of the memory span tests could be generalized beyond the visual mode of presentation, an auditory form of IDS (209) and DDS (209) was used. This test and the procedure of administration were identical in every respect to the visual form described above, with the one exception that the presentation of the stimuli was auditory rather than visual. In the Delay condition the pluses and minuses were also auditory. The auditory version was made by reading aloud (a woman's voice) the visual version of the test into a tape recorder as the programmed sequence was automatically presented. This exactly preserved all the time relationships of the programmed visual test. In the auditory version the speaker of the tape recorder occupied the same position as the display screen in the visual form. The tapes, etc. were shielded from the *S*'s view.

The two equivalent forms of the auditory test were identical to their visual counterparts.

2. *Immediate Color-Form Span (ICFS, 2-9) and Delayed Color-Form Span (DCFS, 2-9)*. - This test is identical in every way to the Digit Span tests described above, except that instead of digits there were nine colored forms: triangles, circles, and squares colored either red, yellow, or blue. Each of the nine colored forms was simply substituted for one of the nine digits of the previous test. The purpose of the Color-Form version was to create greater intralist similarity and thus greater interference. A red triangle and a red square, or a blue circle and a yellow circle, for example, would be expected to be more readily confuseable than, say, the digits 3 and 7 or any other possible pairs of digits. For every item in the Color-Form series there are two other items

of the same shape and two other items of the same color with which to confuse the target item. While it is true that each Color-Form must be specified by two items of information (i.e. shape and color), the test is actually more difficult than a digit series twice as long. The reason for this is presumed to be the greater intralist interference of the Color-Form series, although some of the greater difficulty could be due to the fact that two items of information per second have to be processed by the *S* during the presentation of the Color-Form series while there is only one item per second in the digit series.

A wall chart showing all the color-forms was within the *S*'s view at all times; each form was labeled with the initials of its color and shape (e.g., red triangle was RT). *Ss* were instructed to record their responses by writing only the initials of the color-forms. No one seemed to experience any difficulty with this aspect of the Color-Form task.

3. *Retroactive and Proactive Inhibition of Digit Span.* - This test consisted of four conditions:

   a. Digit span (DS) with retroactive inhibition (RI) for digit series lengths of from 4 to 7 digits, thus labeled as *DS-RI, 4-7.* In this paradigm the *S* viewed one series of digits, then a second series, and then was asked to recall the first series.

   b. Digit span with proactive inhibition (PI) for series of from 4 to 7 digits: *DS-PI, 4-7.* This paradigm is formally the same as the one above, except that the *S* is asked to recall the second series.

   c. Delayed Digit Span (*DDS, 4-7*). The *S* views a digit series followed by an interval filled with a series of red pluses and minuses and then has to recall the digit series. This is the *control* condition or baseline for the RI and PI paradigms.

   d. *Immediate Digit Span* (IDS, 4-7). In this condition the digit series is recalled without delay.

The paradigms of all four of these conditions were explained to the *S* with the aid of a wall chart showing the sequence of events for each condition. The *S* would never know until the time of recall whether he would have to recall the first series (RI condition) or the second (PI). The signal for recalling the first series was one bong; the signal for recalling the second series was two bongs.

In all conditions (except IDS or immediate recall) 13 seconds always intervened between the last digit of the series-to-be-recalled and the bong which signaled recall. This interval was filled with other digits or with the "+" and "-" signs or with both.

Since the RI and PI conditions the first and second series could each have anywhere from four to seven digits, all possible combinations were used, making 16 RI and 16 PI series. The DDS and IDS conditions each had 16 series. Thus there were 64 recall series in the whole test.

In all cases the two series were temporarily separated for three seconds during which a blue disc (2 inches in diameter) appeared in the center of the display screen. The sequence of events for each of the four conditions was as follows.

RI Condition (First digit series to be recalled):

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
first digit series (4 to 7)	4-7 seconds
blue disc	3 sec.
second digit series (4 to 7)	4-7 sec.
+ and -	8 minus 2nd series
blank	1 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

PI Condition (second digit series to be recalled):

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
first digit series (4 to 7)	4-7 sec.
blue disc	3 sec.
second digit series (4 to 7)	4-7 sec.
+ and -	11 sec.
bong	1 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

Control Condition (Delayed Digit Span):

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
digit series (4 to 7)	4-7 sec.
+ and -	11 sec.
blank	1 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

"Immediate" Digit Span Condition:

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
digit series (4 to 7)	4-7 sec.
+ or -	1 sec.
blank	1 sec.
bong (signal to write)	1 sec.
blank (for writing response)	13 sec.
bong (ready signal)	1 sec.
etc.	

It should be noted that in the above "immediate" digit span condition there was actually a delay of three seconds before the *S* wrote his response. Also note that in all four conditions the *S* had to use the "x" and "-" pushbuttons, if only to press one of them once at the end of a series (as in the "immediate" condition). Again, it should be noted that in all of the conditions except the "immediate" there was an interval of 13 seconds between the end of the digit series to be recalled and the time of recall. Thus the RI and PI conditions were perfectly equated in terms of the total time lapse between the end of the recall series and the time of recall, and both were equated in this respect with the "control" condition.

4. *Forward and Backward Digit Span*. - This test consisted of four conditions, with 16 replications of each. The conditions were as follows:

a. *Forward Digit Span Pre-cued (FDS-Pre)*. In this condition the digit series was both *preceded* and *followed* by a cue informing the S that the order of recall was to be in the forward direction.

b. Forward Digit Span Post-cued (FDS-Post). In this condition the *S* was given no knowledge before or during the presentation of the series concerning the order of recall; the signal for recalling the series in the forward direction was given only *after* the series was presented.

c. *Backward Digit Span Pre-cued (BDS-Pre)*. The cue for the recall of the series in the reverse of the order of presentation was given both *before* and *after* the presentation.

d. *Backward Digit Span Post-Cued (BDS-Post)*. The cue for recall in the backward direction was given only *after* the presentation of the series.

All series in this test consisted of 15 digits, but the *S* was asked to recall no more than 12, and the answer sheets provided only 12 "boxes" for recording each series. The reason for this procedure was to "force" the *S* to recall the series in the cued direction. Recall of the three digits at the beginning of the series would, therefore, never be credited if the *S* was supposed to recall the series in the backward direction. And recall of the three digits at the end of the series would never be credited if the *S* was supposed to recall the series in the forward direction. This procedure also has the effect of reducing the "recency effect" in the case of forward recall and of reducing the "primacy effect" in the case of backward recall.

These points were made clear to the *S*s in the preliminary instructions, which were given with the aid of a wall chart that depicted the sequence of events in each of the four paradigms. In recording their recall on the answer sheets, *S*s were instructed to write the digits in the row of boxes in the order of left to right for the forward span and in the order of right to left for the backward span. All *S*s were given a practice test consisting of two instances of each of the four conditions.

There were eight pages of answer sheets; each page contained eight rows of boxes. To minimize confusing the *S* concerning the experimental conditions at any one time, each of the pre-cued conditions was confined to a separate page of the answer sheets. The order of the various conditions was as follows:

Page 1 - FDS-Pre

Page 2 - BDS-Pre

Page 3 - FDS-Post and BDS-Post

Page 4 - FDS-Post and BDS-Post

Pages 5-8 - same as Pages 1-4

In the post-cued condition, of course, the order of the forward and backward conditions was always random within each set of eight series.

The pre or post cue for recall in the forward direction was the appearance of a red "+" sign on the display screen. A red "-" sign was the cue for backward recall. *S*s were required to press the "+" and "-" pushbuttons whenever the corresponding cue signals appeared on the screen; this helped to insure that *S*s were paying attention to these cues.

The sequence of events in the four conditions was as follows:

Conditions

FDS-Pre	FDS-Post	BDS-Pre	BDS-Post	Time
bong (ready)	bong	bong	bong	1 sec.
+	blank	-	blank	1 sec.
series (15)	series	series	series	15 sec.
+	+	-	-	1 sec.
bong (write)	bong	bong	bong	1 sec.
blank	blank	blank	blank	13 sec.
bong (ready)	bong	bong	bong	1 sec.
etc.	etc	etc.	etc.	

The rules governing the composition of the digit series were somewhat different from those applied to other tests, due to the fact that there were 15 digits in each series of the Forward-Backward digit span test. The order of the digits in each series was random , with five restrictions: (*a*) all nine digits appeared in every series, (*b*) no digit was repeated more than once, (*c*) repeated digits were never temporally adjacent in the order of presentation, (*d*) in each 15-digit series three of the repeated digits always occurred in either the last three positions (in the case of forward recall), or in the first three positions (in the case of backward recall), so that the recall series of 12 digits never contained more nor less than three repeated digits, and (*e*) adjacent digits were never in normal ordinal sequence in either forward or backward recall.

5. Digit Span (DS, 5-12) and Repeated Digit Span (RDS, 5-12). - This test was intended to measure the increment in memory span resulting from a single repetition of the presentation of the digit series before the *S* was required to recall it. There were two conditions: (*a*) recall after a single presentation of the series, and (*b*) recall after two presentations of the same series. Series lengths varied from five to twelve digits. There were five replications of each condition. Thus, in all,

the S was required to recall 80 digit series during the test session.

The single and repeated series were randomly interspersed within sets of eight series each. The S was never informed beforehand whether a series was to be repeated. If the presentation of a series was to be repeated, it was followed by a blue disc, which appeared for one second before the repeated presentation of the digit series.

Series of nine digits or less were made up according to the same rules used in the previous tests. Series of ten to twelve digits contained all nine digits plus three repeated digits; all the other previously stated rules of series composition prevailed.

The sequence of events was as follows.

Digit Span (DS, 5-12):

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
digit series (5-12)	5-12 sec.
+ or - (signal to write)	1 sec.
blank (for writing)	13 sec.
bong (ready signal)	1 sec
etc.	

Repeated Digit Span (RDS, 5-12):

Events	Time
bong (ready signal)	1 sec.
blank	1 sec.
digit series (5-12)	5-12 sec.
blue disc	1 sec.
repeat of digit series (5-12)	5-12 sec.
+ or - (signal to write)	1 sec.
blank (for writing)	13 sec.
bong (ready signal)	1 sec.
etc.	

It should be noted that the red plus or minus sign served as the signal to write, and the S was required to press the corresponding pushbutton whenever the sign appeared. There was never a bong at the end of a series. Ss were instructed to keep their hands poised over the pushbuttons during the presentation of the series.

E. Interference in Serial Learning

Two types of serial learning tasks were used to assess interference effects. One set of tasks involved intra-list interference. The other test involved the retroactive inhibition (RI) of a first-learned list by a second-learned list.

2-2

Certain aspects of procedure were common to all tasks. The usual anticipation method of serial learning was used throughout. The rate of stimulus presentation was always three seconds, with an intertrial interval of six seconds. The first item of the serial list was always preceded by a red light of three seconds duration which served as the stimulus for anticipating the first item in the list. All lists were composed of nine items. The reason for using nine-item lists presented at a three-second rate was that we already had a great deal of previous experience with such lists presented under these conditions and therefore had a good idea of how many lists Ss could reasonably be expected to learn in a single test session and what the range of individual differences was apt to be in our college population This obviated the need for much piloting with the serial learning procedures.

In all conditions the S learned the serial list to a criterion of mastery, i.e., one errorless trial.

In all the serial learning experiments the experimenter (E) sat several feet behing and slightly to the left of the S. E recorded on a specially prepared form the S's performance throughout the course of learning.

1. Serial Intralist Interference (SII). - This test consisted of several parts, which took two one-hour sessions to administer The two sessions were never given on the same day and were usually one week apart.

The basic idea of the test was to contrast the S's learning of serial lists having a low degree of intralist interference with those having a low degree of intralist interference with those having a high degree of intralist interference. The lists of low intralist interference were composed of specially selected trigrams; the lists of high intralist interference were composed of colored forms There were three lists of each type As a preliminary practice test the S was required to learn a list of owrds under exactly the same procedural conditions that applied to the other lists. The order of administration was as follows:

First Session:	Second Session
Practice Test (word list)	
Trigram List A	Color-Form List B
Color-Form List A	Trigram List C
Trigram List B	Color-Form List C

Selection of the words and trigrams was based on the Underwood and Schulz frequence count (Underwood & Schulz, 1960, pp. 336-369). The words and trigrams have approximately the same mean frequency of occurrence in English. The word list consisted of: END, JOY, MAN, HAT, WIT, PIE, SUM, GAS, TIP. The trigram lists consisted of various permutations of: COM, REL, ZIN, VEN, EST, FAC, TIS, MUL, PIM.

The color-form lists consisted of triangles, circles, and squares colored red, yellow, and blue.

The rules for permuting the items in the three forms of each type of list were: List pairs A and B and B and C did not have any items in the same serial position or any pair of adjacent items in the same sequential order. In the color-form lists adjacent items were never of the same shape or color.

The instructions to the S for learning the word and trigram lists were as follows:

- a. On the screen you will be shown a series of words, presented one at a time at the rate of one every three seconds The same order is repeated over and over.
- b. Your task is to learn this series, by anticipating each word before it appears.
- c. The series is preceded by a red light, which is the signal to make the first anticipation.
- d. The first time through the list, simply read each of the words. The next time through you are to anticipate each word before it appears. It is all right to guess when you are not sure.
- e. The series is repeated until you are able to anticipate every word correctly throughout the entire list.

After the first list, the S was simply told he would learn the next list in the same manner. College students seem to take to serial learning quite readily and conform to the instructions in a remarkably uniform manner.

2. Serial Retroactive Interference (SRI). - This test consisted of two parts administered in two sessions. The sessions were always at least one day apart. The two parts of the test were (a) retroactive inhibition (RI) from interpolated learning (IL) of low interference (IL-Low) and (b) RI from IL of high interference (IL-High). In both cases the original learning (OL) and the relearning (RL) were of a 9-item color-form list. The low interference interpolated list (IL-Low) consisted of trigrams (a permutation of the same list used in the intralist serial interference task described above). The high interference interpolated list (IL-High) consisted of a permutation of the color-form OL list. This is a condition known to produce maximal retroactive interference The permuted color-form list in IL never had any item in the same serial position as the OL list and no pair of adjacent items in the OL list were adjacent in the IL list.

OL, IL, and RL were administered in almost immediate succession, with about one minute elapsing between lists. OL, IL-Low, RL were administered in the seccond session. All lists were learned to a criterion of mastery.

The instructions for learning the color-form lists were read to the S only before the presentation of the first list, which were as follows:

- a. On this screen you will see a series of nine colored forms. They consist of triangles, squares, and circles colored red, yellow, and blue. The same shape or color will never occur adjacently in the series For example, a red triangle would never be followed by anything red or triangular.
- b. The first item in the series is always preceded by this red light (E points to red light at bottom of display screen).

c. The items are presented one at a time at a 3-second rate.
d. Your task is to learn the order of the items in the series, by anticipating each item before it appears on the screen.
e. The first time through the series I will name each item as it appears.
f. The next time you see this red light go on, you should try to anticipate the next item, and when it appears, anticipate the item after it, and so on. It is all right to guess when you are not sure.
g. The series will be repeated in this manner until you are able to anticipate every item in the series correctly.

F. Interference from Response Competition: The Stroop Test

The Stroop Test yields, among other things, a direct measure of interference due to response competition. The Stroop was taken by every S in this entire investigation.

As a part of this project a comprehensive review was made of the literature on the Stroop Test. Since this review is very extensive, being at least as long as this entire report, it is being published separately (Jensen & Rohwer, in press).

There is no standardized version of the Stroop, but all variations on Stroop's original procedure (Stroop, 1935) have the following in common: there are three cards - the color card (Card C) on which there are 100 patches of from three to five different colors, the word card (Card W) on which are printed (in black and white) the names of the colors, and the color-word card (Card CW) on which are printed the names of the colors, but printed in an ink of a conflicting color (e.g., the word RED might be printed in green, yellow, or blue, but never in red ink). Each card has 100 items to be named. The S's task on Card C is simply to utter the names of the colored patches as rapidly as possible, scanning the rows from left to right. On Card W the S reads aloud the color names as rapidly as possible. On Card CW the S is required to name the colors of the inks while ignoring the conflicting printed color names The Ss basic score on each card is the total time (in seconds) he takes to utter the 100 names.

Materials. The cards were made large enough to be used as wall charts, so that a S standing four feet from the chart would have no visual difficulty in discriminating the colors or printed words. It had been found previously that when small cards were used, which the S would hold in his hands, it was impossible to control the S's distance from the card, the angle of viewing, or to prevent the S's pointing with the finger or other variations in behavior which interfered with standardized administration. The plates, which were made photographically, were each 17 3/4 in X 25 1/4 in. and were mounted on heavy (1/8 in. thick) cardboards 19 in. X 26 in.

Card C consisted of ten rows and ten columns of evenly spaced colored dots against a flat-black background. The dots were 5/8 in. in diameter and their centers were separated by 1 1/2 in. All the colors were vivid and easily discriminable: red, green, orange, blue, and yellow. The order of the colors was random except for the following restrictions: (a) all five colors appeared an equal number of times, (b) all colors appeared in each row of ten dots, and (c) adjacent dots (in the order in which they were read, from left to right) were never of the same color.

Card W consisted of twenty rows and five columns of the printed names of the colors. The words were in white against a flat black background. All letters were block caps 5/16 in. high. Their line width was 1/16 in. The words were distinctly separated, with the rows and columns exactly in line. The names were in a random order except for the same restrictions that applied to Card C (above). However, the order of the color names on Card W was never congruous with the order of the colored dots on Card C.

Card CW consisted of the same plate as was used for Card W, but with the words vividly tinted with the five colors, the actual color, of course, always conflicting with the color name The order of the colors was the same as their order on Card C.

A Pretest Card was used at the beginning of the test to aid in explaining the task to the S. It consisted of the same plate as Card C, but without any colors on the dots, which were all white on a black background.

Administration. The cards were placed on an easel at about the S's eye level. The S stood facing the easel at a distance of approximately four feet. The Pretest Card was shown first and S was told he would next be shown a similar card with colored dots and that his task was to name the colors, going from left to right, as rapidly as possible, without stopping until the end. The five colors were named by the experimenter (E). Then Card C was presented, E said "Go!" and simultaneously started a stopwatch. E tapped the card with a pencil whenever S made an overt error This procedure seemed to militate against careless performance and overt errors were very rare. The procedure was similar for Cards W and CW. Prior to the presentation of each card, S was told what was expected of him. On Card W, S was told to read the color names as fast as possible, and on Card CW he was told to name the colors and ignore the printed words.

The whole test takes about eight to ten minutes.

G. Personality Inventories

When all sorts of personality questionnaires are factor analyzed at the item level, the two largest super-factors to emerge may be identified as extraversion (E) and neuroticism (N). These two factors account for most of the variance in the personality domain (Eysenck, 1960). The inventories that have been devised specifically to measure these factors in more or less pure form are the Maudsley Personality Inventory (MPI) and the Eysenck Personality Inventory (EPI). All Ss in the present study were given the MPI. The 50 Ss in the All-Tests Group had the MPI and also the two equivalent forms of the EPI.

The MPI has been described in detail elsewhere (Eysenck, 1959; Jensen, 1958). In brief, it consists of two scales of 24 items each, one for Extraversion (E Scale) and one for Neuroticism (N Scale). The EPI is essentially a new version of the MPI and measures the same factors, E and N. Whereas in the MPI there is generally a slight correlation between the E and N factors, this correlation has been entirely eliminated in the EPI. The EPI has two equivalent forms. Further details concerning the construction of these inventories and their theoretical rationale are to be found in the manual of the EPI (Eysenck & Eysenck, 1964).

The 50 Ss in the All-Tests Group filled out the MPI and the two forms of the EPI in the laboratory at the end of the test sessions. Ss in the Single-Test Groups were given the MPI at the conclusion of the first laboratory session and were told to fill it out on their own time and return it when they came back for the second session. This method of administering the inventory worked very well; the return rate was virtually 100 percent and statistical analysis has shown that the results obtained by this manner of administration are strictly comparable to those obtained when the Ss fill out the inventory in the laboratory.

H. Ancillary Data

Scores on the Terman Concept Mastery Test, a high-level measure of intelligence, as well as the college Grade Point Average, were obtained on most of the Ss in the Single-Test Group (N = approximately 300). These data, however, were not used in the present investigation, though they are being analyzed in relation to the experimental tests and will be reported separately.

Raven's Progressive Matrices Test (Raven, 1958), a measure of general intelligence, and the college Grade Point Average were obtained on all Ss in the All-Tests Group; these data are analyzed in the present study.

Sex and age (in months) were recorded for all Ss.

I. Order of Administering Tests

In the Single-Tests Groups the Stroop Test was administered at the beginning of each session, followed by one of the equivalent forms of one of the learning tasks.

Ten separate laboratory sessions were needed for the All-Tests Group. The sessions were at least one day apart and all ten sessions took place within a two week period for each subject. Each of the ten sessions began with the administration of the Stroop Test. The other tests were administered in the following order.

Session	Test
1	Immediate Digit Span (IDS, 2-9) and Delayed Digit Span (DDS, 2-9).
2	Immediate Color-Form Span (ICFS, 2-9) and Delayed Color-Form Span (DCFS, 2-9).
3	Serial Intralist Interference (SII). Practice Test, Trigram List A, Color-Form List A, Trigram List B.
4	Serial Intralist Interference (SII). Color-Form List B, Trigram List C, Color-Form List C. Eysenck Personality Inventory (Form A).
5	Digit Span (DS, 5-12) and Repeated Digit Span (RDS, 5-12).
6	Retroactive and Proactive Inhibition of Digit Span (DS-RI, 4-7; DS-PI, 4-7; DDS, 4-7; IDS, 4-7).
7	Serial Retroactive Interference, Low Interference Condition. Eysenck Personality Inventory (Form B).
8	Serial Retroactive Interference, High Interference Condition. Maudsley Personality Inventory.
9	Forward and Backward Digit Span, Pre and Post Cued.
10	Raven's *Progressive Matrices* Test (45 minutes time limit).

V. Analysis and Interpretation of the Data

The reader will be aided by a word about the organization of this section. It has four main divisions: A. Memory Span Tasks; B. Serial Learning Tasks; C. Stroop Test; D. Factor Analysis of All Tests. Though personality measures were used in connection with every test, discussion of the personality results is included entirely in Section D. The first three sections concern the characteristics of the various tests used for measuring learning and interference effects. In these cases the results are based on the Single Test Groups, i.e., those groups whose main purpose was to provide information concerning the adequacy of each of the tests for the aims of the present study. None of these Single Test Groups took more than one of the experimental tasks. All, however, took the Stroop Test and the Maudsley Personality Inventory.

To facilitate finding the descriptions of the tests and to obviate repeating this information, the tests are referred to here by the same numbered headings under which they are described in the preceding section on Procedures.

A. Memory Span Tests

Scoring Methods. - There are innumerable methods of scoring a *S*'s performance on memory span tests. Forty years ago Guilford and Dallenbach (1925) reported 27 different scoring procedures in the literature; they added a twenty-eighth method and more undoubtedly have been added since then. Most of these variations have little to recommend them, and many at face value are patently redundant. It seemed important for the present study, however, to consider using two main types of scoring system, since there was the possibility that each of these types of scores might involve a psychologically different aspect of learning or encoding a digit series. The two kinds of scores might, therefore, interact with IDs and the results obtained with one method of scoring might not hold true for the other method. Therefore, both methods of scoring were applied to all of the memory span data, and both sets of scores were subjected to practically all the analyses in order to determine whether our conclusions were in any way a function of a specific scoring method. These two main methods of scoring are called Position Scores and Sequence Scores.

Position Scoring consists of giving one point credit for every item recalled in the serial position it occupied in the order of presentation. Thus, if a S recalled the series 12345 as 23451, his Position Score would be zero, whereas if he recalled it as 13425 he would have a Position Score of 2. Since 13425 appears in some ways to have retained less of the information in the original series than 23451, it seems unfair to assign this performance a score of zero. Furthermore, it might be the case that some Ss encode sequential information in terms of absolute position in the series, while others retain associations between adjacent items. Position Scoring would favor the first type of learner and handicap the second type.

Sequential Scoring, therefore, seemed a logical necessity to supplement Position Scoring. Sequential Scoring consists of giving one point credit for every pair of recalled items that maintain the same adjacency that existed in the presented series. Thus if the series 12345 were recalled as 23451, the Sequence Score would be 3, and if it had been recalled as 13425 it would be scored zero. When recall is perfect, the Sequence Score will, of course, always be one point less than the Position Score, since there are N-1 adjacencies in a list of length N.

There was no attempt to make a composite of the Position and Sequence Scores, for if they yielded different results this fact would only be obscured by combining them.

A S's score (either Position or Sequence) on a memory span task comprising a number of series of different lengths and a number or replications of each length was simply the total number of points gained in all series. Now a question arises concerning the desirability of differentially weighting points in terms of the length of series in which they were gained. The only reason for doing this would be toincrease the sensitivity of the total score to the kinds of effects we are interested in assessing, viz., interference effects. Interference is presumed to result in errors in recall of a digit series, and greater susceptibility to interference effects should result in the occurrence of errors in series of such length that errors occur only rarely, if ever, in the absence of experimentally imposed interference. Therefore it might be possible to achieve a larger separation between scores under interference and non-interference conditions if series were weighted inversely to the probability of their being incorrectly recalled under noninterference conditions. For example, a series of 3 digits almost never results in a recall error under noninterference conditions, while a series of 9 digits is frequently missed even under noninterference conditions The S's recall was therefore weighted as follows: The number of items correct in any series was divided by the maximum possible score, i.e. the number of items in the series (multiplied by 100 to remove the decimal). Thus, if the series had 3 items and the S recalled 2 correctly, this would contribute 2/3 or .67 to his total score; and if the series had 9 items and the S recalled 8 correctly, this would contribute 8/9 or .89 to his score. It can be seen that, by this method of weighting, an error in a short (easy) series counts for more in the S's total score than an error in a long (difficult) series. Scores yielded by this method of weighting are called Percentage Scores; and they can be derived either from Position Scores or Sequence Scores.

These total Percentage Scores were obtained in all sets of the memory span data, for both Position and Sequence scores. It turns out, however, that we can dispense with this complication of dealing with Percentage Scores right at the outset. This system of weighting proved to be of no advantage whatsoever, and, in fact, had surprisingly little effect on the rank order of S's total scores. The correlations between the weighted and unweighted (raw) scores were all well over .90 and were always higher than any of the reliability estimates for the scores. Also, the reliabilities of the Percentage Scores did not differ significantly or systematically from the reliabilities of the raw scores. Since the Percentage Scores appear to be completely redundant, no further use or mention of th[illegible]ill be made in the subsequent analyses.

As we shall see, the Sequence Scores turn out to be almost equally redundant and we could have done just as well using nothing more than the raw Position Scores, but there are a few minor complications in this matter which deserve more detailed comment later on. For the time being, therefore, the Sequence Scores will be retained in our analyses.

A Rate of Decrement Index was devised as another means of possibly obtaining a sensitive measure of interference effects as they are manifested in memory span performance. It is known that the percentage of the items recalled decreases at an accelerated rate as the length of the stimulus series increases. It is as if each item added to the series produces an ever increasing interference decrement. The Decrement Index is an attempt to quantify this phenomenon. Here is an example of how the Decrement Index is computed: The S has been presented a number of series of, say, 6, 7, 8, and 9 items in length. The mean percentages of correct responses in each of these lengths of series are, say, 90, 80, 70, 60, respectively. The Decrement Index consists of the sum of the successive differences between these mean percentages, divided by the sum of the mean percentages (X 100, to remove the decimal). Thus:

Series:	6	7	8	9
Mean Percent Recall:	90	80	70	60

Successive Differences: 10 + 10 + 10 = 30

Total of Percents: 90 + 80 + 70 + 60 = 300

Decrement Index: $\frac{30}{300}$ X 100 = 10.00

It will be noted that the Decrement Index is mathematically independent of the S's absolute level of performance. It reflects only the average rate by which each additional item in the series decreases the S's ability to recall the items in the series. A S whose performance deteriorates markedly as the series get longer will obviously have a greater Decrement Index than a S whose recall is not so markedly depressed by increases in series length.

1. Immediate Digit Span and Delayed Digit Span (IDS, 2 - 9 and DDS, 2 - 9).

a. Serial Position Effects in Recall. - Table 1A shows the mean probability of recall of each item in the series as a function of its serial position, the length of the series, and whether the recall was immediate (I) or delayed (D). (The decimal points in the probabilities are omitted.) The p values are based on 1000 observations derived from 100 Ss with 10 replications per S. Table 1B shows the amount of decrement in probability of recall due to the interposed delay interval. It should be remembered that the delay interval was filled by the S's pushing buttons labeled plus and minus in accordance with "+" and "-" signs presented on the screen at a 1 sec. rate. Thus the delay constitutes an interpolated interference, and the decrement in response probabilities is considered to be an interference effect. Differences greater than .080 between the p values in these Tables may be regarded as significant beyond the .01 level. Since the data in these Tables summarize the most important features of S's performance on this task, a number of points in the Tables should be especially noted:

i. The mean proportion of correct recalls ($\bar{X}_p$) decreases at an accelerating rate as the series increase in length. The difference between the I and D conditions (Table 1B) first increases as the series increase in length from 2 to 7 digits, then the difference becomes less for series lengths 8 and 9. This suggests that beyond about 7 digits there is so much intralist interference from various sources even in the I condition that the extra interference in the D condition cannot add much more to the total interference effect. In other words, even in the I condition the S's performance is determined more and more by interference effects as the series increase in length.

ii. The D condition affects recall of series of every length and at every serial position. The overall mean decrement in p value due to delayed recall is .074. There is a definite serial position effect for every series; the low point in the serial position curves is practically always the next to the last position in the series; this is true for both I and D conditions.

iii. By going down the columns in Table IA one can note the decrement in recall of items which have the same serial position but which are followed by increasing numbers of succeeding interfering items. Also note that the D condition results in increasing decrements at all positions up to a particular length of series, after which the decrements decrease (Table IB). The 7-item series is most affected by delayed recall.

Table 1A

Probability of Recall in Visual Digit Series (IDS, 2-9 and DDS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each p Value Based on 1000 Observations (100 Ss X 10 Replications)

Series Length	Time of Recall	Serial Position									
		1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	I	982	987								985
	D	975	962								968
3	I	983	981	985							983
	D	981	965	951							966
4	I	982	979	977	975						978
	D	954	937	919	833						911
5	I	976	956	940	949	961					956
	D	932	850	830	794	826					846
6	I	982	936	907	890	906	921				924
	D	971	899	843	800	772	794				847
7	I	942	907	861	795	779	749	786			831
	D	916	775	691	644	627	558	519			676
8	I	896	778	724	662	594	599	550	557		670
	D	872	764	683	604	536	499	428	444		604
9	I	914	728	623	605	552	505	494	418	458	589
	D	856	733	597	523	464	396	359	311	340	508

Table 1B

Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 1A), as a Function of Length of Series and Serial Position (Visual Presentation)

Series Length	Serial Position									
	1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	007	025								016
3	002	016	034							017
4	028	042	058	142						068
5	044	106	110	155	135					110
6	011	037	064	090	134	127				077
7	026	132	170	151	152	191	267			155
8	024	014	041	058	058	100	122	113		066
9	058	-005	026	082	088	109	135	107	108	080

iv. If the *S*'s memory span is defined as the length of series which he recalls perfectly 50 percent of the time, a corollary is that the *S*'s span is the length of series in which the weakest item's *p* value is .5. The probability of recalling the whole series correctly can be no higher than the probability of recalling the most difficult item. We see that for our present sample of *Ss*, the memory span, as conventionally defined, must lie somewhere between 8 and 9 items for condition I and between 7 and 8 items for condition D. The nadir of memory span would be the length of series in which the most difficult item had a probability of recall not greater than chance. The lower limit of chance can be regarded as $p = .111$ in the present test, and it can be seen that even in the 9-item list under the D condition the most difficult item (Position 8) is well above the chance level. Longer series would be needed to tap the full range of *Ss*' capacities for recalling digit series.

It might be of interest to compare the recall of the 9-item series as presented under the conditions of this experiment with the recall resulting from *simultaneous* (rather than successive) presentation of similar 9-item digit series to similar *Ss*[1]. Each series was presented for 3.7 sec. There was immediate written recall. The *p* values are based on 2432 observations.

---

[1] These data were made available to the writer by Dr. Arthur W. Melton.

Position:	1	2	3	4	5	6	7	8	9	$\bar{X}_p$
p Value:	977	929	926	887	811	819	702	656	798	834

The mean *p* value for simultaneous presentation exceeds that for successive presentation by .245. The most difficult position in the simultaneous series has a *p* value of .656, which means that by this method of presentation a 9-item series is on the average well within the *Ss'* memory span. Thus it seems clear that successive presentation results in greater interference and poorer recall than simultaneous presentation. There are probably two main reasons for this: (*a*) the traces of each item have a longer time to decay and to be interfered with in successive presentation, and (*b*) successive presentation, especially when the *S* has no prior knowledge of the length of the series, does not lend itself so readily to grouping of the digits or other encoding techniques which may facilitate retention. Another factor might be that there was actually more time for rehearsal of the whole series in the simultaneous condition than in the successive condition. The successive condition, to be sure, allowed more rehearsal time per individual item, but the crucial aspect of rehearsal might be the time available for rehearsing the series as a whole.

b. *Effects of Interference and Delay in Recall*. - Another way of viewing these data is to compare the mean number of items recalled for every length of series under both immediate (I) and delayed (D) recall conditions (Position Scoring).

Length of Series:	2	3	4	5	6	7	8	9
Immediate Recall:	1.98	2.96	3.93	4.83	5.49	5.78	5.26	5.30
Delayed Recall:	1.94	2.93	3.66	4.35	5.15	4.90	4.97	4.49

It will be noted that the greatest absolute recall is achieved in the 7-digit series for condition I and in the 6-digit series for condition D. After this peak is reached, further increases in the length of the series result in less absolute recall, which is evidence that succeeding items interfere with the preceding items. The apparent discontinuity at 7 (or any other number of digits) has no special significance. The fact that absolute recall, as a function of series length, increases up to a point and then decreases only reflects the more general fact that the proportion of items recalled decreases, as a function of series length, at a rate greater than the rate of increase in series length. It should be emphasized that each digit beyond the first one produces interference; the interference does not begin to come into play until after the series length exceeds the *S*'s conventional memory span. Thus it must be the case that the *S*'s memory span is in some degree a function of his susceptibility to interference. The delay in recall, filled with interfering activity, only *adds* to the burden of interference that acts even in immediate recall.

c. *Effects of Practice*. - Since there were five replications of every digit series on each of two days (10 replications in all), it is possible to examine the effects of practice on performance. The results are shown below in terms of the mean proportion of digits recalled in all series (Position Scoring). Thus, if a *S* recalled every series correctly, the proportion recalled would be the total number recalled (2+3+ . . . +8+9 = 44) divided by the maximum possible recall (44), which is 44/44 = 1.00. Each proportion is based on 50 *Ss*.

Day 1	Mean Proportion Recalled	
Replications	Immediate	Delayed
1	.72	.59
2	.75	.69
3	.77	.69
4	.82	.74
5	.82	.73
Day 2		
1	.81	.71
2	.81	.77
3	.84	.73
4	.83	.85
5	.84	.80

A trend of improvement with practice is evident in both the Immediate and Delayed recall conditions. Analysis of variance shows that the differences between replications are highly significant (*F* - 33.15, *df* - 4/196, *P*<.001). Thus there is a small but true improvement in performance on this test as a result of practice, and the interaction between replications and conditions of recall (Immediate *vs*. Delayed) shows that improvement with practice is significantly greater for the Delayed Recall condition (*F* = 11.84, *df* = 4/196, *P*<.001). It would be interesting to know the precise mechanisms underlying this practice effect in memory span tests. It is evident from the above data that, for one thing, *Ss* learn to become less susceptible to the interference in delayed recall, and the effects of practice last for a period of at least one week, which was the interval between Day 1 and Day 2.

d. *Correlation Between Scoring Methods*. - All the results reported so far are based on Position Scores. The Sequence Scores give an almost perfectly parallel picture, which should not be surprising when we look at the correlations between the two kinds of scores. For *Ss*' total recall scores, the correlations (Pearson *r*) between Position and Sequence Scores are .97 and .98 for the Immediate and Delay conditions, respectively. Thus, for these tests one of the scoring systems is quite redundant. Therefore, although all the subsequent analyses were carried out on both sets of scores, only the results for Position Scores will be reported here. In all cases, of course, the analyses based on the two kinds of scores have been compared to make sure that there are no discrepancies in the results or conclusions due to the method of scoring.

e. Descriptive Statistics of Ss' Performance.- The means and standard deviations (SD) of the variables included in the correlational analyses are listed below. For these data N = 50.

	Day 1		Day 2	
Measure	Mean	SD	Mean	SD
Total Recall:				
Immediate (I)	172.22	22.25	182.94	20.69
Delayed (D)	152.66	30.16	171.26	25.41
Difference (I-D)	19.56	17.32	11.68	17.64
Relative Difference $\frac{I-D}{I} \times 100$	11.38	10.78	6.10	9.29
Decrement Index:				
Immediate	5.40	2.93	5.80	2.87
Delayed	7.76	3.87	8.00	3.77

A number of points should be noted in these data:

i. All the differences between the Immediate and Delayed conditions are highly significant. Analysis of variance shows the overall I-D difference to be significant at well beyond the .001 level. The experimental variable clearly had the intended effect.

ii. The difference between Days 1 and 2 is significant beyond the .001 level, again indicating a significant practice effect on this task. Since this practice effect (as reflected by Day 2 - Day 1) may be regarded as a measure of learning or learning-to-learn, it is of interest to note whether there are significant IDs in this measure. Analysis of variance shows that there are significant IDs in improvement with practice. The Ss X Days interaction is significant ($F = 3.65$, $df = 49/196$, $P < .001$). The reliability of the Day 2 - Day 1 difference scores is .73. The triple interaction (Ss X Days X Condition of Recall), however, is not significant ($F < 1$). No use was made of this measure of overall practice effect. IDs in practice effects have not been differentiated from true error variance in the interpretation of the present data. In interpreting the factor analysis of the All-Tests Group, which took only the Day 1 forms of each test, there was no feasible way of taking account of IDs in practice effects, which are therefore treated as just one more ingredient in the amalgam of uncontrolled effects that contribute to unreliability of measurement. In the present tests the reliability of measurement is still sufficiently high, despite IDs in practice effects, so there is no real problem. Where day-to-day practice effects contribute more to the ID variance than is contributed by the experimental variables, there is a serious problem. This problem was overcome in the present study simply by eliminating tasks which had this feature.

iii. Note that the Delayed Recall condition adds to the ID variance. The *SDs* of the Delayed Recall scores are significantly ($P < .05$) greater than those of the Immediate Recall scores. Delay increases IDs in performance. Nearly all *Ss* recall less after delay. Only two of the 50 *Ss* performed slightly better under delayed than under immediate recall.

f. *Analysis of Variance*. - A complete analysis of variance was carried out, including all sources of variability: Conditions (I vs. D), Days (1 vs. 2), Replications (5), Series Lengths (2-9), *Ss* ($N = 50$), and all of the possible interactions among these variables. The complete analysis is presented in Appendix A. The analysis was used mainly to test the significance of the experimental effect and its interactions with other variables and to determine the reliability of various measures derived from the data.

The analysis shows that true error variance, in the sense of error of measurement, constitutes only 6.6 percent of the total variance in these massive data. IDs among *Ss* contribute 50.3 percent of the variance. The remaining 43.1 percent of the variance is, of course, contributed by the experimental variables (Conditions, Days, Replications, Lengths) and their interactions. It is apparent that most of the variability is attributable to IDs.

g. *Reliability of Scores*. - Reliability can be determined in two ways: from the analysis of variance (Appendix A) or from the correlation (Pearson r) between two sets of scores. Both methods have been applied here. Also, two kinds of reliabilities can be determined: reliability based on the correlations between Days 1 and 2, and reliability based on the intercorrelations among the Replications. We will designate these two kinds of reliability as r_D (based on correlation between days) and r_R (correlation between replications). The reliability estimate always pertains to a single form or administration of the test; that is, it represents the average reliability of Day 1 or Day 2. The reliability of the composite of Day 1 + Day 2 would have to be obtained by using the Spearman-Brown formula, boosting the test length by a factor of 2. The reason for the present procedure of estimating reliability for only a single administration of each test is that the reliabilities are to be used in the interpretation of the total intercorrelation matrix of the All-Tests Group, which was given only one form of each test.

The reliabilities of various measures are given below. Only those reliabilities obtainable from the analysis of variance in Appendix A are given here. Reliabilities of all scores, however, were obtained by means of Pearson r. The two methods of estimating reliability show no appreciable difference.

Measure	Reliability Estimate Analysis of Variance		Pearson r
	r_D	r_R	r_D
Total Recall (I+D)	.76	.92	.76
Immediate Recall (I)			.70
Delayed Recall (D)			.79
Difference (I-D)	.70	.61	.69
Relative Difference (I-D)/I			.69
Decrement Index			
Immediate			.54
Delayed			.41

h. *Correlational Analysis of Scores*. - The analysis of variance has shown that the interaction between *Ss* and conditions of recall is highly significant (F=4.09, df = 49/196, P < .001). This is reflected also in the substantial reliability of the I-D difference score. These facts rule out the possibility that D = I-k, where k is a constant for all *Ss*. The possibility that D = kI is ruled out by the fact that the Relative Difference, (I-D)/I also has substantial reliability. In other words, the retention loss due to delay is not a constant proportion of what is retained in immediate recall. Thus we conclude from these two facts that the IDs in the effects of delayed recall are both significant and authentic.

The following intercorrelations are based on Days 1 and 2 combined. The correlations obtained within Days 1 and 2 separately were not significantly different. Other variables were included in the correlation matrix (Decrement Index, Stropp Scores, Neuroticism, Extraversion, sex, and age), but these are not germane to the present discussion and will be taken up separately in later sections.

Variables	2	3	4
1. Recall I	.79	(-.06)	(-.19)
2. Recall D		(-.66)	(-.74)
3. I-D			.99
4. (I-D)/I			

Only two of these correlations are worth considering, r_{12} and r_{34}. r_{12} indicates that I and D have a good deal of their variance in common (about 70 percent after correction for attenuation) but do not overlap completely. Thus it appears that at least two factors are involved in these two measures, which is simply a further form of confirmation of the existence of IDs arising from delayed recall. What it is that the I and D scores have in common and what it is that they do not share cannot be known at this stage of analysis. It might be worth while, however, to keep the following possibilities in mind as hypotheses. Assume that two or more factors are involved in the total true variance of the I and D scores. Then we can hypothesize that (*a*) the largest part of the variance in both tasks, the 70 percent that I and D share in common, is a memory span factor and the 30 percent they do not share represents an interference factor existing only in Recall D; or (*b*) the IDs variance shared by I and D scores consists both of learning and of intratask interference factors and that what they do not share in common is a third factor contained only in the D scores and consisting either of a different kind of interference than that operating in Recall I or of recall decrement due to sheer delay in recall and thus bringing in a factor of rate of decay of the memory trace; or (*c*) these two or three factors (learning, interference, and rate of trace decay) are all operating in both Recall I and Recall D, but have different weights in the two tasks, interference and decay factors contributing somewhat more to the variance in Recall D than in Recall I. This third alternative is the most likely and most consistent with the kind of conclusions generally arrived at in factor analysis. That at least two factors are operating in IDs in memory span is almost a certainty, but that is as far as we can legitimately go in our conclusions at this point.

The correlation of .99 between the two types of difference scores (r_{34}) shows they are measuring the same thing and again reflects the fact that we are dealing with true IDs in delay decrement, or at least with IDs that cannot be attributed to some constant difference or ratio between I and D scores.

The correlations in parentheses should be ignored. Like practically all correlations between gain (or decrement) scores and the components from which they are derived, they almost certainly involve a spurious element which rules out any possiblity of meaningful interpretation.

1A. Auditory Form of the Immediate and Delayed Memory Span Tests. The procedures and analyses of the auditory test completely parallel those of the visual form of the test described above. Our main concern here is with the differences in our findings for visual and auditory tests of memory span.

a. Serial Position Effects in Recall. - Table 2A shows the mean probability of recall as a function of series length, serial position, and condition of recall (Immediate or Delayed). Table 2B shows the decrements in recall due to delay (I-D). The overall picture is very similar to that of the visual tests (Tables 1A and 1B). The main (and significant) differences between the auditory and visual forms of the tests are in the relative effects of Conditions I and D. The overall mean p values for I and D on the visual tests are .864 and .791; the corresponding values for the auditory tests are .873 and .718. In other words, with immediate recall the auditory test is easier than the visual, whereas the reverse is true for delayed recall. The interpolated "pluses" and "minuses" (which were also auditory) in the delay condition apparently are much more interfering in the auditory than in the visual form of the test. A possible interpretation of this finding is that information is encoded in the STM store in the form of auditory images, whether the input is visual or aural. This should make auditory presentation easier but also more susceptible to interference in the auditory modality. Also, Ss can probably respond to the visual "+" and "-" signs without encoding them in any way that would interfere much with STM, while hearing the words "plus" and "minus" could more readily interfere with STM traces if these traces are encoded as auditory images.

b. Effects of Interference and Delay in Recall. - Below is shown the mean number of auditory digits recalled as a function of length of series, and the differences between auditory (A) and visual (V) presentation. (Auditory data based on 50 Ss X 10 replications.)

Length of Series:	2	3	4	5	6	7	8	9
Immediate Recall:	2.00	3.00	3.78	4.93	5.57	6.08	5.67	5.47
Delayed Recall:	1.96	2.90	3.64	3.92	4.40	3.90	3.67	3.93
A-V (Immediate):	.02	.04	-.15	.10	.08	.30	.41	.17
A-V (Delayed) :	.02	-.03	-.02	-.43	-.75	-1.00	-1.30	-.56

The A-V differences are considerably greater in the delayed recall condition than in the immediate recall. An important aspect of these findings is that they indicate that the delay condition involves more than the passage of time

Table 2A

Probability of Recall in Auditory Digit Series Tests (IDS, 2-9 and DDS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each p Value Based on 750 Observations (50 Ss X 10 Replications + 50 Ss x 5 Replications)

Series Length	Time of Recall	Serial Position 1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	I	997	997								997
	D	984	973								979
3	I	998	997	998							998
	D	981	954	957							964
4	I	996	982	918	929						956
	D	976	926	878	818						900
5	I	997	978	962	969	983					978
	D	917	796	744	672	689					764
6	I	986	932	876	844	896	950				914
	D	940	805	698	628	568	636				712
7	I	980	922	870	805	761	750	858			849
	D	893	700	594	466	428	340	373			542
8	I	921	794	708	639	592	588	615	700		695
	D	817	631	517	459	376	320	255	272		456
9	I	900	733	588	589	469	465	473	508	670	599
	D	830	710	509	409	361	284	247	227	257	426

Table 2B

Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 2A), as a Function of Length of Series and Serial Position (Auditory Presentation)

Series Length	Serial Position 1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	013	024								018
3	017	043	041							034
4	020	056	040	111						057
5	080	182	218	297	294					214
6	046	127	178	216	328	314				202
7	087	222	276	339	333	410	485			307
8	104	163	191	180	216	268	360	428		239
9	070	023	079	180	108	181	226	281	413	173

during which memory traces decay; the delay involves active interference as well. Simply to say that auditory traces decay more rapidly than visual traces does not seem plausible in view of the superiority of auditory to visual recall in Condition I.

c. Effects of Practice. - This is shown below in terms of the mean proportion of digits in all series recalled (by 50 Ss) on each replication of each day of practice (Position Scores).

		Mean Proportion Recalled	
	Replications	Immediate	Delayed
Day 1	1	.81	.58
	2	.80	.63
	3	.81	.61
	4	.82	.61
	5	.84	.65
Day 2	1	.84	.64
	2	.79	.67
	3	.83	.62
	4	.85	.68
	5	.85	.69

The improvement with practice is significant ($P < .001$) and is manifested mainly in the delayed recall. The effect is not essentially different from that for visual presentation.

d. Correlation Between Scoring Methods. - The correlations (Pearson r) between Position and Sequence scores for Recall I and Recall D are .97 and .97, respectively. Thus, we may conclude that the Sequence score is redundant, and it will not be reported in the subsequent analyses.

e. Descriptive Statistics of S's Performance. -

	Day 1		Day 2	
Measure	Mean	SD	Mean	SD
Total Recall:				
Immediate (I)	180.37	20.05	184.39	20.02
Delayed (D)	136.00	25.68	146.29	28.32
Difference (I-D)	44.37	19.44	38.10	21.48
Relative Diff. (I-D)/I X 100	24.20	10.52	20.31	11.44
Decrement Index:				
Immediate	5.43	3.22	5.00	3.28
Delayed	10.08	3.34	8.76	3.82

The same points made in connection with the visual test apply here as well. In fact, the auditory version of the test exaggerates all the differences between the immediate and delayed conditions.

f. Analysis of Variance. - The same analysis of variance was performed on the auditory as on the visual data. The complete analysis is presented in Appendix A. The analysis shows that error of measurement constitutes only 7.8 percent of the total variance in these data, which is just slightly more error than existed for the visual presentation. IDs among Ss contribute 50.2 percent of the variance, and the remaining 42 percent of the variance is due to the effects of the experimental variables and their interactions.

g. Reliability of Scores. - Reliabilities were determined by exactly the same methods as for the visual tests.

	Reliability Estimate		
Measure	Analysis of Variance		Pearson r
	r_D	r_R	r_D
Total Recall (I+D)	.85	.90	.85
Immediate Recall (I)			.82
Delayed Recall (D)			.80
Difference (I-D)	.61	.70	.62
Relative Diff. (I-D)/I			.62
Decrement Index:			
Immediate			.67
Delayed			.61

The reliabilities of the auditory test are somewhat higher than of the visual test.

h. Correlational Analysis of Scores. - The correlation between Recall I and Recall D (for Days 1 and 2 combined) is .70, and the absolute and relative difference scores (I-D and (I-D)/I) correlate .97. These correlations are essentially the same as those for the visual test.

i. Correlation Between Auditory and Visual Performance. - The crucial question, of course, is whether there are IDs in memory span as a function of sensory mode of presentation. If such an interaction between Ss and modality of presentation existed, it would mean that the conclusions based on all the visual memory span tests would be specific to visual learning and could not be assumed to hold true for the auditory mode as well.

The best way to answer this question is to obtain correlations between the visual and auditory tests and to correct these correlations for attenuation. If the corrected correlations differ significantly from 1.00, it could not be claimed that visual and auditory tests were equivalent with respect to the measurement of IDs.

Two groups of Ss (N=25 in each group) were given both the auditory and visual memory span tests (IDS and DDS, 2-9). One group had the visual test on Day 1 and the auditory test on Day 2 (Group VA). The other group (Group AV) had the auditory test on Day 1 and the visual test on Day 2. Correlations were obtained between Days 1 and 2 for each group and these were corrected for attenuation by the correlations between Days 1 and 2 obtained from the groups which had taken either visual or auditory tests on both Days (Groups VV and AA). The correction for attenuation consisted of $r_{AV}/\sqrt{r_{AA} \cdot r_{VV}}$ and $r_{VA}/\sqrt{r_{AA} \cdot r_{VV}}$. The results are presented below.

Correlations

Measure	Raw r_{AV}	Raw r_{VA}	Corrected r_{AV}	Corrected r_{VA}
Recall I	.86	.78	1.13	1.03
Recall D	.79	.79	.99	1.00

When the correlations are corrected for attenuation they do not differ significantly from 1.00. The same analysis was applied to Sequence Scores and all of the corrected correlations ranged between 1.05 and 1.16. Also, the correlation between Position Scores and Sequence Scores for the auditory data is of the same magnitude as that for the visual data, viz. r = .98. Thus, there is a perfect correlation between IDs in visual and auditory presentation. It is interesting that even though the modality of presentation results in certain significant differences in absolute level of performance under I and D conditions, there is absolutely no Ss X Modality interaction, as shown by the perfect correlation between performance under the two types of presentation. This being the case, the remaining memory span tests were given only by visual presentation.

The perfect correlation between visual and auditory memory span belies the generality of Wechsler's statement that "It is certainly true that a good auditory memory does not go with a good visual memory" (Wechsler, 1958, p. 131). The high correlation between auditory and visual span may, however, by a condition that exists only in normal Ss. Certain types of brain damage (e.g. temporal lobe lesions) reportedly result in great discrepancies between visual and auditory memory span. Also, the perfect correlation between auditory and visual memory span may not hold up in all age groups (e.g., McGhie, Chapman, & Lawson, 1965).

2. Immediate and Delayed Color-Form Span (ICFS and DCFS, 2-9).

It will be recalled that the Color-Form memory span test parallels the Digit Span test in every way except that nine colored geometric forms (triangles, squares, and circles colored red, yellow, and blue) were used instead of the nine digits.

a. Serial Position Effects in Recall. - Table 3A shows the proportion of items recalled as a function of serial position, etc. Table 3B shows the mean difference between the I and D conditions. It is immediately apparent that the color-form test is much more difficult than the digit series test. The mean p value for Recall I is only .490 and for Recall D it is .375, as compared with .864 and .791 for the visual digit span tests. Also, delayed recall results in greater loss in the color-form test.

The color-form series may be regarded as involving two items of information (color and form) for every item. Its greater difficulty could be due to this fact, at least in part. But intralist interference is presumed to contribute more to its difficulty. One line of evidence that amount of information is probably not the only factor involved is obtained by comparing the average recall of color-form series with digit series twice the length of the color-form series. Color-form series of lengths 2, 3, and 4 (Recall I) have *p* values of .946, .826, and .618. The *p* values for digit series of lengths 4, 6, and 8 are .978, .924, and .670. Thus, color-form series are even more difficult than digit series of twice the length. It would be interesting to know how color-form *p* values would compare with those of digit series twice the length and presented at a rate twice as fast. We have no information on this point.

Table 3A

Probability of Recall in Color-Form Series (ICFS, 2-9) and DCFS, 2-9) as a Function of Length of Series, Serial Position, and Time of Recall (Immediate vs. Delayed). Each *p* Value Based on 550 Observations (50 *Ss* X 5 Replications + 30 *Ss* X 10 Replications)

Series Length	Time of Recall	Serial Position									
		1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	I	947	944								946
	D	860	829								844
3	I	865	805	807							826
	D	684	544	491							573
4	I	811	680	514	465						618
	D	680	493	349	220						436
5	I	709	649	385	262	274					456
	D	587	453	378	162	113					339
6	I	680	540	442	209	122	153				358
	D	582	407	274	162	058	067				258
7	I	742	569	387	238	082	065	118			314
	D	531	373	262	125	034	038	056			203
8	I	658	516	371	138	045	040	049	093		239
	D	514	416	307	138	040	034	024	033		188
9	I	531	354	227	138	053	024	022	040	107	166
	D	538	409	245	109	042	020	007	022	054	161

Table 3B

Amount of Decrement in Probability of Recall Due to Delay of Recall (i.e., the Difference I-D in Table 3A), as a Function of Length of Series and Serial Position (Color-Form Series)

Series Length	Serial Position 1	2	3	4	5	6	7	8	9	$\bar{X}_p$
2	087	115								101
3	181	261	316							253
4	131	187	165	245						182
5	122	196	007	100	161					117
6	098	133	168	047	064	086				099
7	211	196	125	113	048	027	062			112
8	144	100	064	000	005	006	025	060		051
9	-007	-055	-018	030	011	004	015	018	053	006

Color-form span lengths of from two to nine items are more than sufficient to tap the entire range of *S*s' memory span ability. Memory span for these materials, as conventionally defined, is between 3 and 4 items. For series of seven or more items, the *p* values for the weaker positions are even less than the lower limit of chance (.111), if chance is what the *S* could get if he drew the colored forms out of a hat at random. This means, of course, that many *S*s were failing to give any responses at all for these positions in the series. Many *S*s, in fact, reported that beyond the first three or four items their minds were a "complete blank," except for the last item in the series, and even this item is not recalled better than chance in series with more than six items. In delayed recall the memory span for colored-forms is somewhere between 2 and 3 items.

Another interesting point, which was observed to a somewhat lesser degree in the digit series, is that delay takes its greatest toll in the region of the conventionally defined memory span and has lesser effects both below and above this point (Table 3B). Delay does not have a significant effect on the 9-item color-form series. Interference is so great even in Recall I that the extra interference added in the delay interval can do no further damage.

b. *Effects of Interference and Delay in Recall.* - Below is shown the mean number of items recalled in every length of series under immediate and delayed recall (Position scoring). (The Means are based on 30 *S*s X 10 replications).

Length of Series:	2	3	4	5	6	7	8	9
Immediate Recall:	1.92	2.55	2.44	2.31	2.11	2.16	1.99	1.48
Delayed Recall:	1.72	1.66	1.65	1.63	1.53	1.41	1.54	1.43

These results suggest that the effects of intralist interference must be enormous for color-form series. The absolute recall for a 9-item series is even less than for a 2-item series, and between series of three and of nine items the *S*s' absolute recall decreases, on the average, by more than 50 percent.

c. *Effects of Practice*. - Each proportion below is based on 30 *S*s.

		Mean Proportion Recalled	
Day 1	*Replications*	*Immediate*	*Delayed*
	1	.29	.17
	2	.31	.25
	3	.36	.25
	4	.39	.24
	5	.36	.29
Day 2			
	1	.41	.28
	2	.41	.33
	3	.38	.31
	4	.45	.34
	5	.44	.35

The practice effect is significant ($P < .001$) as shown by the main effect for replications in the analysis of variance. The degree of practice effect is about the same for immediate and delayed recall, whereas digits showed a significantly greater practice effect for delayed than for immediate recall. The mean difference between Days 1 and 2 is the same for both forms of recall in the color-form test. One wonders if this practice effect is due to the *S*s' learning to encode color-forms as single units rather than as items each containing two units of information. At present, however, we have no clue concerning the mechanism of practice effects on this test. *S*s may develop strategies for overcoming interference, such as "turning off" their attention after a certain number of items have been presented, to protect them from interference of subsequent items in the series. Future experiments can be designed specifically to test hypotheses about the mechanism of practaice effects in memory span. All that the present study definitely shows is that highly reliable practice effects do in fact occur.

d. *Correlation Between Scoring Methods*. - The correlations between Position and Sequence scores for Immediate and Delayed recall are .94 and .98, respectively. The Sequence scores are therefore regarded as redundant and are not reported in the following analyses.

e. *Descriptive Statistics of Ss' Performance*. - These data are based on *N*=30.

Measure	Day 1 Mean	Day 1 SD	Day 2 Mean	Day 2 SD
Total Recall:				
Immediate (I)	76.50	15.50	92.87	20.97
Delayed (D)	53.53	19.24	71.90	22.23
Difference (I-D)	22.97	11.21	20.97	11.89
Relative Diff. (I-D)/I X 100	30.73	15.10	22.67	13.89
Decrement Index:				
Immediate	22.83	5.79	18.10	4.47
Delayed	27.77	8.68	22.57	7.29

Though mean recall is much lower under delayed than under immediate recall, the *SDs* are larger under delayed recall. However, not one of the 30 *Ss* performed better under delayed than under immediate recall. *Ss* differ only in the degree to which they are affected by delay and not in the direction of the effect.

f. *Analysis of Variance*. - The complete analysis of variance of the color-form test is presented in Appendix B. Error of measurement constitutes 10.4 percent of the total variance; IDs contribute 70.4 percent, and the remaining 19.2 percent is attributable to the experimental variables and their interactions.

g. *Reliability of Scores*. - The reliabilities were determined in the same manner as for the digit span tests.

Measure	Reliability Estimate Analysis of Variance r_D	Reliability Estimate Analysis of Variance r_R	Pearson *r* r_D
Total Recall (I+D)	.81	.92	.83
Immediate Recall (I)			.69
Delayed Recall (D)			.89
Difference (I-D)	.44	.34	.44
Relative Diff. (I-D)/I			.66
Decrement Index			
Immediate			.64
Delayed			.65

There is one important difference in the reliabilities of the digit and color-form tests. The I-D score is much lower for the color-form data, despite the high reliabilities of the I and D scores separately. This is a result of I and D being highly correlated. The reason they are more highly correlated than the I and D scores for digits is probably that both the I and D color-form tests are heavily loaded with interference; IDs in both are largely determined by the same factor, which is probably susceptibility to interference.

h. Correlational Analysis of Scores. - In the analysis of variance the Ss X Conditions interaction, when tested against Ss X Condition X Replications, gives an F = 2.01, df - 29/116, $p < .01$. Therefore there are significant IDs in the effects of delay, but these IDs are much smaller for the color-form than for the digit series. The correlation of .88 between Recall I and Recall D (for Days 1 and 2 combined) shows that the two measures have a great deal in common. The fact that the absolute and relative difference scores have significant, though rather low, reliabilities, however, means that I and D do not have all of their true variance in common. The fact that the color-form test seems to be heavily influenced by interference and that the extra interference involved in Recall D adds relatively little to the overall difficulty of the task tends to support the hypothesis that memory span involves both learning and interference and that the interference is probably of the same type as that imposed during the delay interval. Different kinds of memory span materials would thus be viewed as merely having different loadings on the learning and interference factors, depending on such independent variables as intralist similarity, length of delay in recall and amount of interpolated stimuli that demand attention, and so forth. The color-form series is probably very highly loaded on this interference factor, and the delay in recall, with its mild interpolated interference, can only slightly increase the loading of the test on interference.

3. Retroactive and Proactive Inhibition of Digit Span (DS-RI, DS-PI, DDS, IDS, 4-7).

It will be recalled that there are four conditions in this test: Retroactive Interference (RI), Proactive Interference (PI), Delayed Recall (D), and Immediate Recall (I). The Delayed Recall may be regarded as the control condition against which the effects of RI and PI can be assessed.

Scoring. - In addition to the scoring methods applied to all the other memory span data, the present data were subjected to one other form of scoring which was intended to detect Ss who had not complied with the instructions for taking these tests. Since the RI and PI conditions consisted of presenting two series (say, A and B) and asking Ss to recall either A (for RI condition) or B (for PI condition), there was the possibility that Ss might not always attempt to recall the list called for by the signal (one bong for the first list, two for the second). This type of error is, of course, of no interest in the present experiment and would only add to experimental error. Therefore a scoring key was made up to coincide with the series not called for by the signal. It was decided that any score on a "wrong" series that was one or more items higher than the mean of the scores on the "right" series of the same length would constitute a failure to comply with instructions. Ss with a total of more than six such failures (in Days 1 and 2 combined) were eliminated from further analyses. There were only two such Ss. When the number of these failures was six or less (out of the possible 64) the Ss were retained, but a correction was applied to the S's score on each single erroneously recalled series. For example, if a S was supposed to have recalled Series A (7-digits) and got a score of zero correct (or any score more than one below his mean for the remaining 7-digit series) and if the S also had a score of, say, 7 correct on Series B (or any number correct more than one greater than the mean of 7-digit series), it was assumed that the *S* had attempted to recall the "wrong" series. In place of the score of zero the *S* would be given the mean of his scores on 7-digit series (or whatever the series length might be) when the correct series had been recalled. This correction was applied to the protocols of 8 *Ss*. RI and PI conditions did not differ appreciably in the frequency of this type of "wrong" list error.

a. Serial Position Effects in Recall. - Table 4A shows the proportion of digits recalled as a function of serial position, series length, and experimental condition. Table 4B shows the differences between the conditions. The D-RI and D-PI can be regarded as absolute measures of RI and PI. I-D, as in previous tests, indicates the effect of delay in recall, with the "+" and "-" interference during this interval.

The PI and RI conditions definitely produce greater recall decrements than mere delay of recall. The means for D-RI, D-PI, and I-D are .270, .297, and .040, respectively. There seems to be no consistent difference between RI and PI as a function of length of series.

Conventional memory span for the RI and PI conditions is between 5 and 6 digits, as compared with delayed recall (between 7 and 8 digits) and immediate recall (between 8 and 9 digits).

Table 4A

Probability of Recall in Proactive and Retroactive Digit Series Tests (DS-RI,4-7; DS-PI,4-7; DDS,4-7; IDS,4-7) as a Function of Length of Series, Serial Position, and Experimental Condition (Retroactive Interference, Proactive Interference, Delayed Recall, Immediate Recall. Each *p* Value Based on 624 Observations (52 *Ss* X Replications + 52 *Ss* X 4 Replications)

Series Length	Experimental Condition	1	2	3	4	5	6	7	$\bar{X}_p$
4	RI	780	747	667	708				725
	PI	732	700	608	591				658
	D	964	963	956	952				959
	I	969	972	972	966				970
5	RI	778	684	648	576	611			659
	PI	763	647	602	523	525			612
	D	946	919	879	868	884			900
	I	978	952	941	928	924			945
6	RI	767	600	494	418	372	455		517
	PI	687	600	561	489	467	501		551
	D	941	890	837	794	763	781		834
	I	952	916	876	871	845	873		889
7	RI	636	559	472	396	345	334	337	440
	PI	658	491	453	334	300	286	359	412
	D	925	837	738	681	659	625	645	731
	I	935	886	822	734	715	670	684	778

Table 4B

Amount of Retroactive Interference (RI) in Terms of Decrement in Probability of Recall (D-RI in Table 4A), Proactive Interference (D-PI in Table 4A), and Decrement Due to Delay in Recall (I-D in Table 4A), as a Function of Series Length and Serial Position

Series Length	Type of Decrement	Serial Position 1	2	3	4	5	6	7	$\bar{X}_p$
4	D-RI	184	216	289	244				233
	D-PI	231	263	348	361				301
	I-D	005	009	016	014				011
5	D-RI	168	234	231	293	274			239
	D-PI	183	272	278	345	359			288
	I-D	032	033	061	059	039			045
6	D-RI	174	291	343	375	390	326		317
	D-PI	254	291	276	305	296	280		283
	I-D	011	025	039	077	083	093		055
7	D-RI	289	278	267	285	314	291	308	290
	D-PI	267	346	286	347	359	339	286	319
	I-D	010	049	084	053	056	045	039	047

b. Effects of RI, PI, and Delay in Recall. - The mean numbers of digits recalled for each length of series under each of the experimental conditions are shown below. The means are based on 52 Ss X 8 replications.

Length of Series:	4	5	6	7
Retroactive Interference:	2.75	3.22	2.99	2.87
Proactive Interference:	2.75	3.21	3.52	3.15
Delayed Recall:	3.80	4.43	4.91	4.90
Immediate Recall:	3.91	4.68	5.29	5.25

In this test each recall series under both RI and PI conditions was interfered with by a series of four to seven items. So we can ask the question: what is the mean recall for series of lengths four to seven as a function of the number of items in the interfering series? The results are shown below in terms of the percentage of items recalled in series of lengths 4, 5, 6, and 7 as a function of the number of items in the interfering series. The results are based on 104 Ss (the Single-Tests and All-Tests Groups combined). Each mean is based on 624 observations.

Condition	Number of Interfering Items 4	5	6	7	Mean
RI	70.6	59.1	54.5	54.1	59.7
PI	65.0	61.4	58.0	54.6	59.8

It is apparent from these figures that performance level decreases as the number of interfering items increases, for both RI and PI conditions. Beyond five interfering items each additional item produces a smaller decrement in performance.

c. Effects of Practice. - Since each replication (within Days) involved interfering series of different lengths, we can assess the effect of practice only by observing the difference in performance between Days 1 and 2. We have already seen in previous tests that there are significant practice effects for Immediate and Delayed digit span, so we are concerned here only with practice effects in RI and PI. Analysis of variance of the RI and PI scores for Days 1 and 2 gives the answer. (The complete analysis is shown in Table 5, Section f.) The overall difference between Days 1 and 2 is significant ($F=23.93$, $df=1/51$, $P<.001$). The interaction of Days X Conditions (RI vs. PI) is not significant ($F<1$). In other words, the Ss' performance improves with practice on both RI and PI, but RI and PI do not differ in the magnitude of the practice effect.

d. Correlation Between Scoring Methods. - The correlations (based on Days 1+2) between Position and Sequence Scores are as follows:

Measure	r
RI	.97
PI	.98
Delay	.98
Immediate	.96

The two types of scores are obviously measuring the same basic variable and therefore only the Position scores are used in the analyses.

e. Descriptive Statistics of Ss' Performance. - The following data are based on $N = 52$.

	Day 1		Day 2	
Measure	Mean	SD	Mean	SD
Total Recall:				
RI	44.82	14.07	49.59	15.69
PI	46.55	13.43	54.49	16.40
Delay (D)	68.53	12.39	75.73	11.36
Immediate (I)	74.22	9.65	78.69	9.33
Decrement Index:				
RI	17.33	13.83	9.69	13.27
PI	8.80	11.75	10.82	11.90
D	9.69	6.95	5.49	5.76
I	7.67	5.97	5.14	5.80

f. Analysis of Variance. - Since the RI and PI scores are the central issue in this test, analysis of variance is shown for these scores alone. Both RI and PI differ significantly from both the D and I scores. Analysis of the D and I scores yields essentially the same picture as found in the previous digit series tests involving D and I conditions of recall.

Table 5

Analysis of Variance of Total Recall Scores in Conditions RI and PI

Source of Variance	df	Mean Squares	F	P<
Conditions (C)	1	562.32	8.04	.001
Days (D)	1	1673.55	23.93	.001
Subjects (Ss)	51	502.47	7.19	.001
C X D	1	38.95	< 1	
C X Ss	51	285.40	4.08	.001
D X Ss	51	148.49		
Residual	51	69.93		
Total	207			

The above analysis speaks for itself. RI and PI differ significantly, recall in RI being poorer than in PI, which means that retroactive inhibition is greater than proactive inhibition in this test. Ss differ significantly in overall performance level. But the most interesting finding, from the standpoint of the present study, is the highly significant interaction between Ss and Conditions, which means that IDs in susceptibility to interference depend partly upon whether the interference is proactive or retroactive. The two types of interference are not equivalent as far as IDs are concerned. Ss tend to be consistently more susceptible to one or the other type of interference.

g. Reliability of Scores. - Reliability estimates are based only on the correlations between Ss' scores on Day 1 and Day 2, since the RI and PI conditions in this test have no true replications within Days; series of each length (from 4 to 7) were interfered with (either proactively or retroactively) by series of different lengths (from 4 to 7), and the variance among these conditions would not constitute a proper estimate of experimental error. The reliability coefficients were derived from both the analysis of variance (Table 5) and from Pearson r between Days 1 and 2. In each case the reliability is given for a single administration of the test (N = 52).

	Reliability Estimate	
Measure	Analysis of Variance	Pearson r
Total Recall:		
RI+PI+D+I	.44	.54
RI		.60
PI		.58
RI-PI	.61	
D		.59
I		.37

The reliabilities of the RI and PI scores, as well as of the difference between them, are quite satisfactory. One point is of particular interest in the above figures, viz. the fact that the reliability of the combined scores (RI+PI+D+I) is much lower than the reliabilities of all but one of the single scores. All the other digit span tests in the present battery have shown the opposite, i.e., the highest reliability is found for the sum of the Ss' scores under the various conditions of the tests. The lower reliability found for RI+PI+D+I suggests that the components of this composite score are not highly intercorrelated.

h. Correlational Analysis of Scores. - Since the relationship between IDs and RI and PI promises to be of considerable interest, the correlations between the various scores in this test are shown here for Days 1 and 2 separately and combined (based on the 52 Ss in the Single-Tests Group). In addition are presented the intercorrelations obtained from the All-Tests Group (N = 50), which had the Day 1 form only.

Table 6

Intercorrelations (Pearson r) Among RI, PI, Immediate, and Delayed Recall Scores for Single-Tests Group (N = 52). (Day 1 correlations above the diagonal, Day 2 below).

Measures	1	2	3	4
1. RI		.35	.57	.63
2. PI	.31		.65	.53
3. D	.54	.61		.83
4. I	.47	.50	.71	

Table 7

Intercorrelations Among Variables in Table 6 for Days 1 and 2 Combined

Measures	1	2	3	4
1. RI		.28	.57	.57
2. PI			.65	.48
3. D				.81
4. I				

Table 8

Intercorrelations Among RI, PI, D, and I for All-Tests Group (N = 50)

Measures	1	2	3	4
1. RI		.28	.48	.37
2. PI			.53	.55
3. D				.65
4. I				

It will be noted that the sets of correlations in Tables 6, 7, and 8 show considerable consistency from one administration to another and from one group of Ss to another. It is also interesting to note that the lowest of all the correlations is that between RI and PI, even though these two variables are the most alike in means and *SDs*. RI apparently has more in common with both Conditions D and I than it has in common with PI.

The most extreme correction for attentuation of the highest of the correlations between RI and PI still leaves only about 35 percent of their variance in common. Much of this common variance is probably attributable to the fact that any digit series involves a certain amount of "intrinsic" interference (e.g., interference of successive presentation and of successive recall). The substantial correlations of both RI and PI with D and I scores indicate that both RI and PI also involve the same factors that exist in D and I.

Since the correlations in Table 7 were obtained from the All-Tests Group and were entered into the factor analysis of all the tests, we can get a good estimate of the lower limit of reliability of these measures by noting their communalities in the factor analysis. Reliability can never be lower than the communality and it is usually much higher. The communalities of RI, PI, D, and I are .71, .67, .77, and .77, respectively. The most probably reason that these communalities are higher than the reliabilities estimated from correlations between Days 1 and 2 is that these correlations are measures of stability over time and are affected by interactions between *Ss* and practice effects. The lower bound of reliability given by the communality, on the other hand, represents internal consistency in one administration of the test.

i. *Chi Square Analysis.* - Another way of observing the IDs in RI and PI is by means of the following contingency table, showing the number of *Ss* (Single Tests Group) whose RI scores exceed their PI scores on both Days 1 and 2, the number whose PI scores exceed their RI scores on both days, and the number of *Ss* who are inconsistent from one day to the other. Four *Ss* who showed no difference between RI and PI on one day or the other were, of course, not entered in the table. The results are shown below. For this contingency table the relationship between Days 1 and 2 (or the consistency with which *Ss* are most strongly affected by one or the other type of interference) is highly significant (chi square = 8.1, *df* = 1, *P* < .01).

		Day 2	
		RI > PI	PI > RI
Day 1	RI > PI	14	6
	PI > RI	8	20

j. *The Question of Ss' Strategy.* - The question may be raised whether IDs in RI and PI arise as a result of particular mnemonic or attentional strategies adopted by *Ss* in an attempt to maximize their scores on either RI or PI. For example, a *S* might decide that it is too difficult to pay attention both to the interfering series and to the series-to-be-recalled and might adopt the strategy of always ignoring either the first or the second series. If the series he has concentrated on is the one he is asked to recall, he will have better recall than if he had paid attention to the other interfering series as well. If the reverse occurs, he will score lower than he would if he had paid attention to the series he had decided to ignore. *S*s with a high RI score and a low PI score would be those whose strategy consisted of ignoring the second series; and *S*s with a high PI score and a low RI score would be those whose strategy consisted of ignoring the first series. It cannot be completely ruled out that this is what *S*s were doing and that this might be what accounts for the low correlation between RI and PI. There are several points of evidence that seem inconsistent with this interpretation,however.

i. In the first place, the reliabilities of RI and PI are too high to suggest that approximately half the *S*s were more or less consistently ignoring and responding at random to one or the other condition. The reliabilities of RI and PI are consistent with those of other digit series where there is no reason to suspect that *S*s adopted the type of strategy we have hypothesized to exist in the RI and PI conditions.

ii. The frequency distributions of scores on RI and PI show no signs of a piling up of *S*s at the low-score end of the scale, as one would expect if a sizeable proportion of *S*s were not paying attention to either the first or the second series in each RI or PI condition. In fact, the large majority of *S*s attain scores well above anything like a chance level on both RI and PI. The argument that *S*s might be inconsistent in their strategies throughout the course of testing, which would tend to obscure the appearance of any strategy, is incompatible with the high reliability of IDs in the RI-PI difference as well as with the high reliabilities of RI and PI.

iii. In this test *S*s' performance on the I and D conditions is not significantly or systematically different from the performance of *S*s on other memory span tests involving only the I and D conditions. This would probably not be true if *S*s who obtained relatively superior PI scores did so because they avoided attending to the first series in the PI paradigm. *S*s had no way of knowing that the single series of the D and I conditions were not the first series of the PI condition. If the strategy of a number of *S*s consisted of not fully attending to the first series, the scores on D and I should therefore be depressed, in comparison with the scores of *S*s in other tests who encounter only the D and I conditions. This is not at all the case.

iv. *S*s who are superior in one condition (RI or PI) are not, on the whole, complete failures on the other condition. This indicates that *S*s were trying to recall all the series they were called upon to recall. Below is shown the mean number of digits recalled per series by *S*s who were either superior or inferior on the RI condition (Single-Tests Group, *N* = 52).

	Mean Recall	
	RI	PI
*S*s with RI > PI:	3.40	2.65
*S*s with PI > RI:	2.65	3.50

The corresponding data for the All-Tests Group (*N* = 50) are shown below.

	Mean Recall	
	RI	PI
*S*s with RI > PI:	3.72	2.88
*S*s with PI > RI:	3.18	3.78

v. Probably the most convincing evidence that IDs in RI and PI are not a result of the type of strategy we have hypothesized comes from the factor analysis of the RI and PI scores along with all the other tests used in this study. As we shall see, RI and PI have quite different patterns of factor loadings. It seems very unlikely that this could come about as a result of Ss' making a decision as to which list (i.e., the first or the second) they would attend to and which list they would ignore. We would expect that the list attended to would correlate highly with the other memory span tests, and it should make no difference whether the list attended to was in the RI or the PI condition. The overall reliability of the RI and PI scores should be lowered by a strategy approach, but there is no reason to expect a different pattern of factor loadings for RI and PI.

Further attempts to interpret this finding of IDs in RI and PI will be postponed until the section dealing with the factor analysis of all the tests.

4. Forward and Backward Digit Span

There were four conditions in this test: Forward (F) and Backward (B) recall; and each series was either Pre or Post cued by a signal instructing the S to recall the list in either the forward or backward direction. All series were 15 digits in length, but Ss were instructed to recall only the first 12 items in either the F or B direction.

a. Serial Position Effects in Recall. - Table 9 shows the mean proportion of items recalled as a function of serial position and experimental condition. Note that serial position in this Table refers to the order of recall and not the order of presentation; in the Forward condition order of presentation and order of recall are, of course, the same, but they are opposite for the Backward condition. The results in Table 9 are based on the Single-Test Group ($N = 30$) combined with the All-Tests Group ($N = 50$).

Table 9

Probability of Recall in Digit Series Tests as a Function of Serial Position, Order of Recall (Forward vs. Backward), and Time S was Informed of Order of Recall (Pre-Cued vs. Post-Cued). Each p Value Based on 1760 Observations (30 Ss X 32 Replications + 50 Ss X 16 Replications).

Condition	Serial Position in Recall												
	1	2	3	4	5	6	7	8	9	10	11	12	$\bar{X}_p$
For. Pre	664	549	513	480	436	429	364	305	218	155	110	110	361
For. Post	470	400	331	346	296	257	186	148	140	110	115	140	245
Back. Pre	698	625	576	500	443	384	251	196	132	079	065	106	338
Back. Post	573	507	437	363	318	233	156	106	075	065	068	072	248

The overall mean *p* values for the F and B conditions are .253 and .293. Note that backward span is superior to forward in this test. This is the reverse of what is found in the forward and backward digit span tests of the Wechsler intelligence scales. It appears that the relative difficulty of F and B spans is a function of the length of series. When the series is relatively short, F will be easier than B; when the series is long, the reverse will be true.

The Forward-Backward test was intended to maximize interference effects. Practically all of the interference is probably due to successive presentation and reproduction. In a long series, the first items suffer a great deal of retroactive interference from succeeding items; the last items of the series suffer less of this retroactive interference. These conditions favor backward recall.

Pre or Post cuing produces larger differencesthan F or B recall. The mean *p* values for the Pre and Post cue conditions are .350 and .247. This cuing variable was introduced to permit the emergence of strategies for minimizing interference effects. It was hypothesized that *Ss* would tend to "turn off" their attention when they subjectively experienced the maximum interference they could tolerate if they were to hold on to the items they had already taken in up to a certain point in the series. *Ss* would learn to estimate the point in the series beyond which, if they continued to attend equally to every successive digit, their attempt to recall the list would be impaired by an excessive burden of interference. *Ss* would then develop techniques of attending that would help to protect memory traces from excessive interference. The results shown in Table 9 are consistent with this hypothesis. If *Ss* were actually behaving in accord with this hypothesized strategy, it appears that they were greatly overestimating the optimal point in the series at which to "turn off" their attention, for the proportion of digits recalled even in the first five positions of the Pre-cued Forward condition is much less than for digit series of from 2 to 9 items (see Table 1A) with delayed recall.

c. *Effects of Practice.* - For this test, the effects of practice are most clearly seen in the descriptive data for Days 1 and 2 (Section *e*) and in the complete analysis of variance (Section *f* and Appendix C). In short, the overall level of performance is significantly ($P < .001$) better for Day 2 than for Day 1. There are also significant ($P < .01$) IDs in practice effects, as shown by the *Ss* X Days interaction.

d. *Correlation Between Scoring Methods*. - This is the one test for which the Position and Sequence scores are not equivalent. Below are given the correlations between Position and Sequence Scores.

	Correlation		
Measure	Day 1	Day 2	Days 1+2
For.-Pre	.27	.52	.40
For.-Post	.78	.64	.73
Back.-Pre	.30	.56	.46
Back.-Post	.68	.65	.67

Since the correlations between scoring methods are much lower than the reliabilities of the scores, it will be necessary to look at both the Position and Sequence scores in all the subsequent analyses. The factor analysis of the All-Tests data will also have to be performed on both types of scores, since these might possibly have different factor compositions.

Why should the two scoring methods differ for this test when they were perfectly correlated in other tests? The following reasons seem likely. First of all, these series are 15 items in length and it might be that much less positional information is encoded for lengthy series and Ss tend instead to learn direct associations between adjacent items. It could also be the case that Ss attempt to encode the digits in the form of a number of groups of digits, and the groups of digits might be more resistant to interference than is the information about their exact positions in the list. This would result in Sequence scores being higher than Position scores and, as we shall see, this is in fact the case. Also note that the correlation between scoring methods is markedly lower for the pre-cued than for the post-cued series. This suggests that Ss' strategies may differ in the degree to which they involve the learning of position or of associations between items. As was argued previously, the pre-cued condition favors the adoption of strategies, and there are probably IDs in these strategies.

e. Descriptive Statistics of Ss' Performance. - Since we wish to compare the Sequence and Position scores, and since the two scoring systems differ in the maximum possible score (Sequence is always one less than Position), the results are presented in Table 10 in terms of the percentage of the maximum possible score, which makes the results for the two scoring methods directly comparable.

Table 10

Means and SDs of Percentage of Maximum Possible Score in the Various Conditions of the Forward and Backward Digit Span Tests. (Each Mean and SD Based on 30 Ss X 16 Replications.)

	Day 1				Day 2			
Conditions	Position		Sequence		Position		Sequence	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
For.-Pre	26.50	15.52	30.47	8.61	31.90	21.29	36.90	12.17
For.-Post	20.33	13.21	21.27	9.97	23.07	16.46	28.33	11.51
Back.-Pre	27.03	16.10	31.03	11.22	32.73	20.34	40.03	12.65
Back.-Post	23.03	14.17	24.87	11.10	25.43	16.67	30.60	11.76
Sum of For.	46.83	27.08	51.74	16.63	54.97	36.70	65.23	21.86
Sum of Back.	50.06	29.34	55.90	21.00	58.17	35.71	70.63	22.14
Sum of Pre	53.53	30.84	61.50	17.66	64.63	40.58	76.93	23.36
Sum of Post	43.36	26.02	46.13	19.36	48.50	31.47	58.93	20.94

f. Analysis of Variance. - To make the overall level of scores comparable for Sequence and Position methods of scoring, the analyses of variance were performed on the percentage of maximum possible score rather than on the raw scores. Raw scores and percentage scores are perfectly correlated, of course, and analysis of variance performed on either raw or percentage scores would yield the same F ratios. Percentage scores, however, permit direct comparison of the Mean Squares for both Sequence and Position scores in the analysis of variance presented in Appendix C.

Our main concern is the interaction between Ss and the experimental variables. These may be summarized as follows:

i. For both Position and Sequence scores the main effect for Ss is highly significant ($P < .001$), whether the error term involves Days or Replications.

ii. For Sequence scores the Ss X Cuing (Pre vs. Post) interaction falls short of significance at the .05 level when the error term consists of the Ss X Cuing X Days interaction, but is significant ($P < .01$) when the error term consists of the Ss X Cuing X Replications interaction. For Position scores, the Ss X Cuing interaction is significant at the .001 level for both error terms.

iii. The Ss X Direction (F vs. B) interaction is significant beyond the .01 level both for Sequence scores and for Position scores when the error term involves either Days or Replications.

iv. The triple interaction of Ss X Cuing X Direction is significant at the .01 level for both types of scores and for error terms involving either Days or Replications. This source of ID variance will naturally play some part in determining the pattern of intercorrelations among the four main scores derived from this test.

g. Reliability of Scores. - Reliabilities were estimated by the Pearson r between Days 1 and 2 and by analysis of variance, using correlations between Days (r_D) and between Replications (r_D). The reliability estimates shown in Table 11 are for a single administration of the test.

Table 11

Reliability of Position and Sequence Scores on the Forward (F) and Backward (B) Digit Span Test Under Pre and Post Cuing for Direction of Recall

	Analysis of Variance Estimates				Pearson r	
	r_D		r_D		r_D	
Measure	Position	Sequence	Position	Sequence	Position	Sequence
F:Pre					.87	.61
F:Post					.84	.61
B:Pre					.92	.75
B:Post					.88	.75
F-B	.49	.60	.61	.88		
Pre-Post	.81	.30	.95	.42		
Grand Total	.89	.84	.96	.92	.93	.79

h. Correlational Analysis of Scores. - Intercorrelations (based on Day 1+2) between the scores are shown below. Correlations for Position scores are above the diagonal; Sequences scores are below the diagonal.

Variable	1	2	3	4
1. Forward:Pre		.87	.92	.79
2. Forward:Post	.70		.85	.82
3. Backward:Pre	.78	.61		.91
4. Backward:Post	.64	.68	.78	

These correlations are very high indeed. They tend to suggest that only one general factor would be needed to account for all the intercorrelations. A principal components analysis of these intercorrelations shows that variables 1 to 4 (above) have loadings on the first principal component (or general factor) of .97, .91, .98, and .90, respectively, for the Position scores, and .89, .83, .91, and .88, respectively, for the Sequence scores. If these loadings were corrected for attenuation, there would be practically no variance left unaccounted for by the general factor. This finding does not seem to be in accord with the conclusions we drew from the analysis of variance. The power and sensitivity of analysis of variance for showing interactions to be statistically significant, even though these interactions may account for only a minute proportion of the variance, is the most likely reason for this illusion of a discrepancy between the analysis of variance and the correlational analysis. Though the proportion of variance not accounted for by the general factor in these tests may be significant, as indicated by the analysis of variance, it is so small as to be practically negligible.

For reasons that are not definitely understood, the same state of affairs does not exist in the All-Tests Group where this test is concerned. This is the only test in the whole battery on which the Single-Test and All-Tests Groups show a marked difference. Here are the correlations for the All-Tests Group ($\underline{N} = 50$). (Position scores above diagonal, Sequence scores below.)

Variable	1	2	3	4
1. Forward:Pre		.58	.33	.32
2. Forward: Post	.65		.24	.30
3. Backward:Pre	.33	.26		.54
4. Backward:Post	.15	.09	.76	

This matrix could not be accounted for by a single general factor. Why is it so different from the matrix of the Single-Test Group? The reliabilities of the scores in the All-Tests Group (estimated from the correlation between replications in the analysis of variance) are at least as high as those of the Single-Test Group. The reliability of the F-B difference is .91 for Position scores and .88 for Sequence scores. The reliability of the Pre-Post difference is .66 for Position scores and .73 for Sequence scores.

The most tenable hypothesis to account for this discrepancy between the Single-Test and All-Tests Groups involves changes in factorial composition of the test as a result of prolonged practice on other memory span tasks. This was the first and only test taken by the Single-Test Group, but it was the last test of the whole battery for the All-Tests Group. The All-Tests Group, therefore, was highly practiced on a variety of forward memory span tests. This practice effect apparently interacted with the various conditions of the Forward-Backward test to such an extent that a single factor cannot account for their intercorrelations. The trends in the Single-Test Group which were detectable in the Ss X conditions interactions in the analysis of variance become clearly apparent in the pattern of correlations among the variables in the All-Tests Group.

In order to permit comparison of the performance of the two Groups on this test, the scores of the All-Tests Group are presented below as the percentage of maximum possible score. These results can be compared directly with those of the Single-Test Group shown in Table 10.

Conditions	Position		Sequence	
	Mean	SD	Mean	SD
For.-Pre	43.42	14.36	39.62	12.99
For.-Post	26.58	10.14	23.80	9.91
Back.-Pre	39.46	13.91	41.22	10.87
Back.-Post	28.96	9.20	35.64	7.95
Sum of For.	70.00	22.36	63.42	20.42
Sum of Back.	68.42	21.72	76.86	16.54
Sum of Pre	82.88	23.08	80.84	19.46
Sum of Post	55.54	14.29	59.44	14.46

The All-Tests Group had an overall higher level of performance than the Single-Test Group. Even the Day 2 performance of the Single-Test Group was considerably below the All-Test Group, which had only a single administration of the test. It is interesting that the All-Tests Group showed its greatest superiority in the Forward-Pre test and did not differ so markedly from the Single-Test Group on the other conditions. This is completely consistent with hypothesis that the All-Test Group differs from the Single-Test Group on this task because of greater practice on pre-cued forward digit span tests. The practice effect did not transfer to any great extent to Backward span or to the Post-cued conditions. Nor do the Sequence scores show a very marked superiority in the All-Tests Group, which suggests that a good deal of the practice effect consists of improved retention of positional information rather than of direct associations between adjacent items.

5. Digit Span and Repeated Digit Span (DS, 5-12, and RDS, 5-12).

This test had two parts: recall of a digit series after one presentation, referred to here as Single DS and recall of a digit series after two presentations, referred to as Repeated Digit Series (RDS). Series were 5 to 12 digits in length. The purpose of the test was to measure IDs in the increment in memory span resulting from a single repetition of the series. The question of interest is whether the increment due to repetition taps some factor independent of memory span and possibly identifiable as a "learning ability" factor, as distinct from a span factor.

a. Serial Position Effects in Recall. - These are shown in Table 12A. Table 12B shows the increment due to repetition as a function of serial position and series length.

Memory span as conventionally measured is somewhere between 8 and 9 digits for one presentation, and between 9 and 10 digits for the repeated series. Repetition improved recall for every length of series and at every serial position where there is not an obvious ceiling effect. In general, the longer the series, the greater the benefit from repetition, which is what we should expect if the learning curve is a negatively accelerated function of number of presentations: easy (short) lists are almost completely learned in a single presentation, while difficult lists may require several repetitions to be learned to the same degree. In this sense memory span tests are one-trial serial learning.

Table 12A

Probability of Recall in Digit Series Tests (DS, 5-12, and RDS, 5-12) as a Function of Length of Series, Serial Position, and Number of Presentations of Series (1 vs. 2). Each p Value Based on 550 Observations (30 Ss X 10 Replications + 50 Ss X 5 Replications).

Series Length	Number of Presentations	Serial Position												$\bar{X}_p$
		1	2	3	4	5	6	7	8	9	10	11	12	
5	1	965	947	929	918	936								939
	2	967	973	958	962	922								958
6	1	929	916	896	896	878	869							897
	2	976	953	940	942	944	871							938
7	1	916	882	862	829	784	756	794						832
	2	942	914	896	887	898	876	851						894
8	1	825	813	758	751	696	607	631	831					739
	2	922	876	845	793	734	687	718	814					799
9	1	811	747	745	673	591	498	409	456	754				632
	2	885	827	769	684	642	596	525	569	798				699
10	1	709	676	656	589	502	424	325	249	371	831			533
	2	798	691	680	622	573	516	474	389	453	693			589
11	1	684	618	567	540	469	376	273	204	194	289	669		444
	2	753	678	596	545	511	484	393	338	280	411	807		527
12	1	722	631	564	480	404	313	267	233	227	211	331	680	422
	2	762	704	618	567	516	456	405	315	305	274	409	693	502

Table 12B

Amount of Increment in Probability of Recall Due to One Repetition of Series (i.e. 2-1 in Table 12A), as a Function of Length of Series and Serial Position

Series Length	Serial Position												$\bar{X}_p$
	1	2	3	4	5	6	7	8	9	10	11	12	
5	002	026	029	044	014								019
6	047	037	044	046	066	002							041
7	026	032	034	058	114	120	057						062
8	097	063	087	042	038	080	087	-017					060
9	074	080	024	011	051	098	116	113	044				067
10	089	015	024	033	071	092	149	140	082	-138			056
11	069	060	029	005	042	108	120	134	086	122	138		083
12	040	073	054	087	112	143	138	082	078	063	078	013	080

It is quite evident from Tables 12A and 12B that repetition adds a significant increment to recall, though the increment is surprisingly not very large. In fact, the mean overall increment in p value is only .080. Repetition falls far short of doubling the span. The probability of recalling a 5-digit series perfectly is about .92; the probability of perfect recall of a 10-digit series is about .25, and one repetition raises this probability only to about .39. This suggests that memory span tasks are heavily loaded with interference, for if this were not the case one would expect Ss to nearly double their span when the series is repeated. It is obvious that the S must lose almost as much as he gains during the second presentation of the series. But there is a gain, and this gain can be legitimately regarded as learning.

b. Effect of Repetition in Recall. - The mean number of digits recalled as a function of series length and number of presentations is shown below. (Based on 30 Ss X 10 replications.)

Number of Presentations	Length of Series 5	6	7	8	9	10	11	12
One:	4.64	5.39	5.84	5.77	5.55	5.27	4.90	4.90
Two:	4.82	5.66	6.22	6.38	6.16	5.90	5.59	5.84
Increment:	.18	.27	.38	.61	.61	.63	.69	.91

c. Effects of Practice. - Below is shown the mean proportion of maximum possible score on each of the ten replications of each condition (N = 30).

	Replications	Mean Proportion Recalled: Single	Repeated
Day 1	1	.45	.55
	2	.56	.63
	3	.60	.67
	4	.62	.67
	5	.64	.70
Day 2	1	.62	.69
	2	.65	.72
	3	.68	.73
	4	.66	.71
	5	.68	.70

As in all the other memory span tests, there is significant improvement with practice. The analysis of variance (Appendix D) shows a highly significant ($P < .001$) difference between Days 1 and 2.

d. Correlation Between Scoring Methods. - The correlation between Position Sequence scores (Days 1 and 2 combined) is .98 for DS and .97 for RDS. The two types of scores are therefore regarded as equivalent and no further use will be made of the Sequence scores.

e. Descriptive Statistics of Ss' Performance. - Based on *N* = 30.

Measure	Day 1		Day 2	
	Mean	SD	Mean	SD
Total Recall:				
Single (S)	196.97	39.08	225.13	47.56
Repeat (R)	221.10	32.15	243.63	47.67
Difference (R-S	24.13	18.91	18.50	24.04
Relative Diff. (R-S)/S	13.67	13.22	8.87	11.53
Decrement Index:				
Single	10.87	4.15	8.67	4.42
Repeat	9.17	2.89	6.80	3.75

f. Analysis of Variance. - The complete analysis of variance for this test is presented in Appendix D. The following points are to be noted:

i. The main effect of repetition is significant well beyond the .001 level. Only three of the 30 *Ss* showed no overall improvement in recall as a result of repetition. Thus there can be no doubt that learning occurs when the series is repeated.

ii. The overall differences between *Ss* are highly significant; when the error term is based on the *Ss* X Days interaction, *F* = 13.29, *df* = 29/29, *P* < .001; and when the error term is based on *Ss* X Replications, *F* = 31.43, *df* - 29/116, *P* < .001.

iii. The crucial issue is the *Ss* X Conditions (Single vs. Repeat) interaction. It falls short of significance at the .05 level when the *Ss* X Days X Conditions interaction is used as the error term (*F* = 1.842, *df* = 29/29). When the *Ss* X Conditions X Replications interaction is used as the error term, the *Ss* X Conditions interaction is significant (*F* - 3.55, *df* = 29/116, *P* < .01). Though this effect is significant by this criterion, it is instructive to note how much of the total variance is accounted for by the *Ss* X Conditions interaction -- exactly 1.01 percent, as contrasted with 20.5 percent for the overall differences between *Ss*. All sources of *Ss* variance (i.e., main effect plus all interactions involving *Ss*) constitute 76 percent of the total variance. The experimental variables and their interactions constitute 14.4 percent, and 9.6 percent is attributable to experimental error.

In the more highly practiced All-Tests Group (*N* - 50) the main effect for *Ss* is highly significant (*F* = 21.0, *df* = 49/196, *P* < .001), but the *Ss* X Conditions interaction is barely significant at the .05 level (*F* - 1.47, *df* = 49/196).

g. Reliability of Scores. - Reliability is given below for a single administration of the test, as estimated by the same methods used in previous tests. *N* = 30.

	Reliability of Estimate		
Measure	Analysis of Variance		Pearson r
	r_D	r_R	r_D
Total Recall (S+R)	.86	.94	.90
Single (S)			.89
Repeated (R)			.83
Difference (R-S)	.30	.56	.31
Relative Diff. (R-S)/S			.36
Decrement Index:			
Single			.53
Repeated			.66

The fact that the difference score and relative difference score both have very low reliability, despite the high reliability of the separate scores, suggests that something approximating $R = kS$ or $R = k + S$ might be the case here. The difference R-S involves little true variance. In this respect the R-S is quite unlike the difference between the Immediate and Delayed recall conditions of previous memory span tests.

h. Correlational Analysis of Scores. - The correlations between Single and Repeat spans for Day 1, Day 2, and Days 1 and 2 combined are .88, .87, and .91, respectively. The correlation in the All-Tests Group is .89. Corrected for attenuation, these correlations come close to unity. We must conclude, therefore, that repetition does not constitute an appreciable source of IDs in memory span performance. Whatever ID variance there is in the increment due to repeated presentation of the series has the same source as IDs in memory span after a single presentation. If the increment is regarded as a form of learning, then memory span itself must be a form of learning.

B. Serial Learning Tasks

Scoring Methods. - Many kinds of scores can be derived from serial learning protocols: number of trials to criterion, number of overt errors, number of omissions, percentage of errors, number of correct responses, number of oscillations, number of forward (anticipatory) errors, number of backward (perseverative) errors, an index of degree of blowing of the serial position curve, an index of skewness of the serial position curve, and so forth. Nineteen such derived scores from a 9-item color-form serial learning task ($N = 60$) were intercorrelated and subjected to a principal components analysis, to determine how many independent dimensions exist among all these derived scores. Five significant dimensions resulted, accounting for approximately 70 percent of the total variability in the matrix. The first component accounted for 30 percent of the variance and is the only component for which the interpretation is completely unambiguous. It represents speed of serial learning. Trials to criterion (one errorless trial) and total errors (all types) had loadings of -.94 and -.92 on this component, and they did not differ appreciably in their loadings on any of the other components. The two measures, therefore, can be regarded as equivalent as far as color-form serial learning is concerned. Since total errors is a more fine-grained measure than trials to criterion, this study has used only total errors as the measure of serial learning ability.

The only other measure used in this study in connection with serial learning is an index of oscillation. It has been hypothesized that oscillation is caused by interference (McGeoch & Irion, 1952, p. 118), and so some measure of oscillation tendency in serial learning seems highly appropriate in the present study.

Oscillation consists of a sequence of successes and failures in anticipating an item on successive trials until the criterion is attained. There would be no oscillation if, after an item was once anticipated correctly, it was never missed on any subsequent trial. It was desired to obtain a measure of oscillation which would be mathematically independent of speed of learning and of error rate. By saying that the oscillaton measure should be mathematically independent of number of errors and of error rate what is meant is that no correlation should exist between the oscillation measure and the other measures in a set of artificial data in which oscillations were determined by some purely random method. In other words, oscillation is measured from a baseline that represents the average amount of oscillation that would occur if only pure randomness prevailed. If we measure oscillation from this baseline, any correlation which might be found to exist in the data between the oscillation index and the speed of learning or error rate must be interpreted as a fact of nature and not as an artifact of the method of measurement.

Three items of information from the individual *S*'s protocol are needed to derive the appropriate Index of Oscillation for a single position in the serial list: (1) the total number of correct anticipations (C) of the item up to, but not including, the criterial trial, (2) the total number of errors (E), i.e. overt errors + omissions, up to, but not including, the criterial trial, (3) the number of runs (R) of correct anticipations (C) and errors (E). A *run* is a succession of *Es* (or C's) which are followed and preceded by C's (or E's) or by neither. For example, take a *S* with the following sequence of E's and C;s for a particular item (the criterion is attained on the 10th trial):

Trial: 1 2 3 4 5 6 7 8 9 10
Response: E E C E C C C E E (C)
Runs: 1 2 3 4 5

Here the number of runs is 5. The Index of Oscillation is obtained by taking the ratio of the actual number of runs (R) to the mean expected number of runs (r) in an infinite number of random sequences, given the particular number of E's and C's. The formula for the expected number of runs (r) is: $r = 1 + 2EC/(E+C)$, where E is the number of errors (overt + omissions) and C is the number of correct anticipations. In the above example, $r = 5.44$. The Index of Oscillation for a single item, then, is simply r/R, and it is multiplied by 100 to remove the decimal. The Index of Oscillation for a given *S* is the mean of the Index for all the single items. (In the present study the total rather than the mean was used, since all lists were of nine items and the mean Index in all cases would be 1/9 of the total.)

The characteristics of this Index of Oscillation have been explored previously in various kinds of serial learning data. It has been found that oscillation tendency is generally greater in slow than in fast learners. On the other hand, there is no simple relationship between the difficulty of a serial list and the amount of oscillation it produces. Oscillation always slows learning, but there are other sources of difficulty than the factors that cause oscillation. It has been found, for example, that a word list will produce more oscillation than nonsense syllables, even though the word list, since it consists of familiar, highly integrated units, is learned much more rapidly than the nonsense syllables. The nonsense syllables, however, are not as subject to extraexperimental interference from word association habits and so they produce somewhat less oscillation. In the total balance of all the factors making for the difficulty or easiness of learning a serial list, it generally turns out that words are easier than nonsense syllables. But difficulty should not be thought of as unidimensional.

To summarize: the two measures used in the analyses of the serial learning data are (*a*) total errors (i.e., overt errors + omissions) and (*b*) the Index of Oscillation.

1. *Serial Intralist Interference.*

This test consisted of seven parts: A Word list, three trigram lists, and three color-form lists. All lists were of nine items and were administered under identical conditions. The main point of interest is the relative difficulty of the trigram and color-form lists, which are presumed to differ markedly in intralist similarity and hence in the amount of intralist interference generated.

a. *Serial Position Effects*. - Table 13 shows the mean serial position curves for each of the conditions. The totals in the next to last row are the mean total errors for each condition. The serial position curves are typical for all these lists. It is of interest to note the degree to which *S*s' serial position curves resemble one another in shape. This can be determined from the average correlation between all possible pairs of serial position curves within one group of *S*s (in the present case this is 435 correlations per list). The correlations are based on the serial position curves represented by percentage of total errors at each position. It can be seen that, in general, the average intercorrelation between serial position curves increases with the difficulty of the task as indicated by the total errors. The color-form lists produce the highest agreement among *S*s, probably because there are fewer idiosyncratic associations among color-forms than exist among words or trigrams. The color-form list, in other words, is more homogeneous and gives rise to fewer of the idiosyncratic associations that tend to create dissimilarities among *S*s' serial position curves.

b. *Effects of Practice*. - The practice effect can be seen clearly in Table 13. The lists were administered in the following order: *Day 1*: Words, Tri-A, CF-A, Tri-B; Day 2: CF-B, Tri-C, CF-C. The improvement with practice is especially great in serial learning. The Word list is actually easier than the trigram list, but the reverse occurs here, because the trigram list (Tri-A) has the benefit of the *S*s' prior practice on the Word list. Also note the great improvement in learning the color-form list; CF-A is almost twice as difficult as CF-C. It might be asked how many lists must be learned

Table 13

Means and SDs of Errors as a Function of Serial Position for Lists Composed of Words (W), Trigrams ($\overline{\text{Tri}}$), and Color-Forms (CF), and the Mean Correlation Between Ss' Serial Position Curves. (Single-Test Group, N = 30)

Serial Position	Words Mean	Words SD	Tri-A Mean	Tri-A SD	Tri-B Mean	Tri-B SD	Tri-C Mean	Tri-C SD	CF-A Mean	CF-A SD	CF-B Mean	CF-B SD	CF-C Mean	CF-C SD
1	3.6	3.6	2.7	3.6	1.2	2.3	0.6	0.9	3.3	2.8	3.4	2.5	1.9	1.9
2	4.1	3.6	3.7	3.7	1.2	1.7	1.6	1.7	6.6	4.5	4.0	2.8	4.7	3.6
3	5.7	5.8	6.4	4.2	3.8	3.0	3.2	2.5	7.6	4.7	4.8	4.3	4.3	3.1
4	7.6	6.3	5.9	4.2	4.7	3.8	4.0	2.9	12.8	6.1	8.0	5.1	5.9	5.0
5	8.7	6.3	8.1	6.0	5.5	5.1	5.6	3.9	14.5	5.9	10.7	6.1	7.2	5.5
6	9.0	6.5	8.8	5.8	6.7	4.7	5.4	3.2	16.3	8.1	11.8	9.4	7.5	5.8
7	9.1	5.8	8.3	4.9	5.3	3.9	6.4	4.3	17.3	8.5	12.1	7.7	9.3	6.5
8	8.0	5.7	7.3	5.5	5.6	3.6	4.6	3.4	18.1	8.4	12.9	7.8	9.1	6.7
9	6.3	4.8	6.1	3.9	4.0	2.8	3.3	3.6	14.2	9.2	10.5	7.3	7.3	6.7
Total	62.07	44.47	57.40	36.04	38.07	27.70	34.73	21.42	110.67	51.11	78.33	46.35	57.20	40.65
Mean r	.32		.45		.46		.31		.70		.66		.53	

before the practice effect levels off. The answer, at least for 9-item color-form lists administered under the same conditions used here, is provided by a previous study (Jensen & Roden, 1963) in which three Ss were required to learn a color-form series (always consisting of a different permutation of the same nine items) every day for 20 days. Improvement in speed of learning took place throughout this period until, by the 20th day, all three Ss could learn the color-form list in one trial; the series was within their memory span. The interaction between S and practice effects, of course, tends to lower the test-retest and equivalent forms reliability of serial learning measures. But despite this potential source of "error", the reliability of these tests is quite good, as we shall see in Section d.

c. Descriptive Statistics of Ss' Performance. - Mean errors and the Index of Oscillation shown in Table 14 for both the Single-Test (N = 30) and the All-Tests (N =50) Groups.

Table 14

Mean Errors and Index of Oscillation*for the Set of Serial Intralist Interference Tests

	Single Test Group (N =30)				All-Test Group (N = 50)			
	Errors		Oscillation		Errors		Oscillation	
Test	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Words	62.07	44.47	7.85	1.41	35.88	31.97	7.21	1.87
Tri-A	57.40	36.04	8.26	1.02	48.78	26.12	8.10	1.10
Tri-B	38.07	27.20	7.28	1.48	27.88	20.15	7.04	1.49
Tri-C	34.73	21.41	7.42	1.65	32.90	23.37	6.79	1.93
CF-A	110.67	51.11	8.50	1.46	83.44	48.39	8.16	1.24
CF-B	78.33	46.35	8.23	1.00	71.88	47.83	7.91	1.48
CF-C	57.20	40.65	8.10	1.12	50.44	24.69	8.26	0.93
Sum of Tri	130.20	76.53	22.97	2.86	109.56	55.54	21.93	3.26
Sum of CF	246.20	113.08	24.83	2.29	205.76	90.15	24.33	2.67
Ratio of CF/Tri	2.32	1.40	1.10	0.16	2.13	1.08	1.13	0.20

*The mean Index of Oscillation = 9.00 for purely random oscillation.

d. Reliability. - Two types of reliability were determined for the serial learning data: (i) internal consistency reliability, based on the intercorrelations among the nine items of the serial list, and (ii) equivalent forms reliability, based on the intercorrelations among the various forms of each list administered at different times.

i. The items in a serial list may be regarded in the same way as items of a test with respect to determining the reliability of the total score from the internal consistency among the items. This reliability measure is obtained from the Between Ss term in the analysis of variance involving Positions X Subjects interaction. The internal consistency reliability for the total score is given by (F-1)/F, where F is the variance ratio for the Between Ss term in the analysis of variance. (The average reliability of a single item in the serial list is given by (F-1)/ F+n-1), where n is the number of items in the list.) The internal consistency reliabilities given here are for the total score. These reliabilities have been determined for overt errors and omissions separately and combined. The reliabilities shown below are for the All-Tests Group (N = 50), which was used for the factor analysis.

Test	Overt Errors	Omissions	All Errors
Words	.74	.96	.95
Tri-A	.53	.92	.92
Tri-B	.60	.94	.93
Tri-C	.66	.91	.91
CF-A	.80	.95	.90
CF-B	.70	.95	.96
CF-C	.61	.90	.92

The internal consistency reliabilities for the Index of Oscillation for Words, Trigrams (mean) and Color-forms (mean) are .60, .45 and .49, respectively. Though these reliabilities are not nearly as high as those of the error scores, they are significant (P < .01) and indicate that there are reliable IDs in oscillation.

ii. The equivalent forms reliability was determined from the average correlation among the three forms of each test. Since the composite of all three forms was used in the factor analysis, the average correlations among the forms was boosted by the Spearman-Brown formula to give the reliability of the composite score for the Trigrams and for the Color-Forms. The reliability of the error score as obtained in this manner is .87 for Trigrams and .75 for Color-forms in the Single-Test Group (N = 30); the corresponding reliabilities in the All-Tests Group (N = 50) are .71 and .60.

The reliability of the Index of Oscillation, obtained in the same manner, is .48 for Trigrams and .25 for Color-Forms in the Single-Test Group, and .47 and .56 in the All-Tests Group.

c. Correlational Analysis of Scores. - Correlations among the various tests and scores (errors and oscillation index) are shown below. Correlations for the Single-Test Group (N = 30) are shown above the diagonal; the All-Tests Group (N = 50) are below the diagonal. (For N = 30, r at the .05 level = .36; at the .01 level r = .46. For N = 50, r at the .05 level = .27; at the .01 level r = .35.)

	Errors			Oscillation		
Variable	1	2	3	4	5	6
1. Words (errors)		.64	.18	.50	.07	-.24
2. Sum of Trigrams (errors)	.50		.27	.35	.20	.03
3. Sum of Color-Forms (errors)	.18	.45		.07	.28	.16
4. Words (oscillation)	.49	.33	.19			
5. Trigrams (oscillation)	.22	.56	.40	.37		
6. Color-Forms (oscillation)	.39	.64	.46	.31	.32	

Only two of the correlations in the Single-Test Group are significant at the .05 level or better, while 12 of the correlations are significant in the All-Tests Group. The low reliability of the Oscillation Index for the Single-Test Group is largely responsible for this difference between the two groups.

Oscillation tendency is positively correlated with errors; after correction for attenuation the two measures have about 80 percent of their variance in common.

The correlations of greatest concern are those between Trigrams and Color-Forms, which are .27 and .45 in the Single-Test and All-Tests Groups, respectively. Corrected for attenuation, these correlations become .34 and .69. Thus it appears that IDs in Trigram and Color-Form serial learning have at most less than 50 percent of their variance in common. Words and Trigrams have much more in common than either has in common with Color-forms. Since CF learning is more difficult and produces significantly more oscillation than trigrams, it is presumed that CF learning involves more intralist interference than do the trigrams.

2. Serial Retroactive Interference (SRI).

This test consisted of two retroactive inhibition (RI) paradigms, one for producing a small amount of RI, the other for producing a large amount. The first is referred to as the Low condition, the second as the High condition. Original learning (OL) in both conditions consisted of a nine-item color-form list. Interpolated learning (IL) consisted of a nine-item list of trigrams in the Low condition and of a nine-item list of color-forms (a permutation of the OL list) in the High condition. Relearning (RL), of course, consisted of the same color-form list as OL.

To determine the test-retest reliability of this paradigm, two Single-Test Groups (each of N = 25) were used, one, called the Low Group, to determine the test-retest reliability of the Low RI condition, the other, called the High Group, to determine the reliability of the High RI condition. Equivalent forms of each paradigm were administered a week apart to the respective groups. The All-Tests Group (N = 50) took both the High and Low paradigms at least one day apart.

a. Serial Position Effects. - The serial position curves of the Serial Retroaction data are not essentially different from those of the Serial Intralist Interference data which were summarized in the previous section. It is of some interest, however, to note the amount of RI as a function of serial position. Percentage RI is determined by the usual formula: Percent RI = 100 X (H-L)/L, where H is the RL score after IL-High and L is the RL score after IL-Low. The S's score in all cases is the number of errors (overt + omissions) made up to the criterion (one errorless trial). This formula was used to determine the percentage RI at each serial position. The realts are shown below for Days 1 and 2 of the Single-Test Group and for the All-Tests Group.

	Serial Position								
Group	1	2	3	4	5	6	7	8	9
Single-Test, Day 1:	158	344	812	217	241	247	159	207	293
Single-Test, Day 2:	150	70	119	135	114	149	152	162	286
All-Tests:	1040	141	500	332	242	268	241	208	159

It should be remembered that Percentage RI is based on a comparison of two groups (High and Low) in the case of the Single-Test data; it is based on one group taking both the High and Low conditions in the case of the All-Tests Group.

The above figures show that the degree of RI produced by these conditions is very pronounced indeed. The Low condition may be regarded as a control condition against which the RI of the High condition can be assessed.

The serial position effect of RI does not seem to show any high degree of regularity, and is certainly nothing like the serial position curve for errors.

b. Effects of Practice. - As can be seen from the means in Table 15, there are marked practice effects in this task. The overall improvement in performance between Days 1 and 2 is significant at the .001 level for OL and IL, but not for RL, which shows a nonsignificant Day 1-Day 2 difference. Does practice affect the percentage of RI? Not much. When percentage of RI is based on errors in RL, it is 240 percent for Day 1 and 180 percent for Day 2. This appears to be a marked decrease in amount of RI. But it should be noted that the Low and High Groups differ markedly in OL, and some of this difference in RI might merely be a reflection of a difference in ability between the two groups. When the means of RL are adjusted in the two groups by means of covariance analysis, using OL as the control variable, the percentage of RI becomes 166 for Day 1 and 180 for Day 2. When the measure of RI is based on errors made in the first RL trial only (which is really a measure of recall and not of learning), the percent RI is 84 for Day 1 and 72 for Day 2. Adjusted by covariance analysis, these percentages become 74 and 73, respectively. Using raw means, one would conclude that practice decreases relative RI; using adjusted means one would conclude that practice has no appreciable effect on RI, or, if anything, somewhat increases RI.

c. Descriptive Statistics of *Ss*' Performance. - First, the recall scores, i.e., the number or errors made in the first trial of RL:

	Day 1		Day 2		Days 1+2	
Condition	Mean	SD	Mean	SD	Mean	SD
Recall - Low:	3.96	2.82	4.56	3.38	8.52	5.47
Recall - High:	7.68	1.74	7.84	1.78	15.52	1.65
Percent RI:	93.94		71.93		82.16	

Analysis of variance shows the difference between High and Low recall scores to be highly significant ($F = 70.72$, $df = 1/48$, $p < .001$). This paradigm clearly produces massive RI. The recall scores were not used in subsequent analyses because their reliability was even lower than that of the relearning scores. The recall score is, after all, a very small sample of the *S*'s behavior and can hardly be expected to have very high reliability for individuals. The test-retest reliability of the recall scores, in fact, was not significantly greater than zero.

The means and standard deviations for the total error scores are shown in Table 15, along with the data on the Index of Oscillation.

Analysis of variance was performed on these data for the Single-Test Groups, with the following results:

i. On CL, the High and Low Groups do not show an overall (i.e., Day 1+2) significant difference (*F* = 1.33, *df* = 1/48).

ii. On IL, the High and Low Groups differ significantly (*F* = 6.59), *df* = 1/48, *P* < .05). This is in accordance with expectation, since the interpolated color-form list is not only more difficult than the trigram list but also suffers from associative interference from CL.

iii. On RL, the High and Low Groups show a highly significant difference (*F* = 192.13, *df* = 1/48, P < .001). Again, we see that this paradigm has produced massive RI.

iv. The High and Low Groups show significantly different scores on the Index of Oscillation in RL (*F* = 19.39, *df* = 1/48, *P* < .001). This is perhaps the best evidence we have that oscillation does reflect interference tendency, as McGeoch and Irion (1952, p. 118) hypothesized. It also supports the contention that the color-form series produces more intralist interference than the trigram series, since in the previous section it was shown that the color-form learning produced significantly more oscillation than trigram learning.

d. Reliability of Scores. - Internal consistency reliability is excellent for all conditions; test-retest (equivalent forms) reliability is disappointing. The test-retest reliabilities shown in Table 16 are based on the correlation between Day 1 and Day 2 for the Single-Test Groups (*N* = 25 in each group); the internal consistency reliabilities (explained in Section 1-*d*) are based on the All-Tests Group (*N* = 50).

Table 15

Means and *SD* of Total Errors and the Index of Oscillation for the Various Lists of the Serial Retroaction Paradigm

	Single-Test Groups (*N* = 50)				All-Tests Group (*N* = 50)	
	Day 1		Day 2			
List	Mean	SD	Mean	SD	Mean	SD
Errors						
CL-Low	64.68	28.11	43.96	20.19	54.52	30.00
CL-High	81.64	36.48	42.76	24.97	53.16	30.46
IL-Low	23.80	21.85	18.36	19.09	19.06	11.73
IL-High	40.32	20.09	30.60	18.74	31.88	16.90
RL-Low	9.00	8.54	9.72	7.66	8.62	9.59
RL-High	30.64	15.26	23.40	11.76	29.88	16.06
Percent RI	240.44[1]	----	140.74[1]	----	392.55[2]	526.85
Oscillation						
CL-Low	8.06	0.90	7.83	1.20	7.65	1.17
CL-High	8.11	2.15	7.70	1.38	7.59	1.40
IL-Low	7.35	2.05	6.27	2.82	6.64	1.95
IL-High	8.05	1.36	7.87	1.56	8.06	1.24
RL-Low	5.33	3.83	5.26	3.19	3.88	2.88
RL-High	7.59	1.58	8.47	1.27	7.68	1.90

[1] Percent RI based on RL means of High and Low Groups.

[2] Percent RI based on mean percent RI for individual *S*s (note the enormous *SD*). The percent RI based on the groups means of the High and Low RL scores is 247.20.

Table 16

Test-Retest (Equivalent Forms) and Internal Consistency Reliabilities of the Various Lists of the Serial Retroaction Paradigm

	Test-Retest		Internal Consistency		
Score	Total Errors	Oscillation	Overt Errors	Omissions	Total
OL-Low	.42	.20	.66	.92	.93
OL-High	.60	-.09	.75	.90	.92
IL-Low	.94	.80	.58	.83	.86
IL-High	.56	.42	.36	.90	.92
RL-Low	.52	.34	.49	.87	.89
RL-High	.32	-.05	.54	.90	.90

The fact that internal consistency reliability is so much higher than test-retest reliability probably means that the factorial composition of serial learning changes with practice. The high internal consistency indicates that any single test is probably a quite reliable measure of *Ss'* performance. But performance seems to have rather little stability from one stage of practice to another. The rather mediocre test-retest reliability cannot be attributed to any marked shrinkage of the variance as a result of practice (see Table 15). Also, the fact that each of these tests has a high communality in the factor analysis means that the tests must be reliable, even though they do not correlate highly between occasions. One could question whether or not we have truly equivalent forms of each list. The lists are quite equivalent in mean difficulty, and it seems unlikely that there should be an interaction between *Ss* and various permutations of the very same items that compose the equivalent forms of each list. The most reasonable hypothesis is that the factorial composition of serial learning changes with practice. This is a disturbing thought in terms of the aims of the present study, because it means the serial learning tests as a whole will not have a great deal of variance in common and will not make for very clear-cut interpretation in the factor analysis. Each serial task will tap a somewhat different conglomerate of factors, depending on the stage of practice. This complicated problem will have to be investigated in its own right in future studies.

e. *Correlational Analysis of Scores*. - The intercorrelations among the errors and oscillation index of the various lists are given in Table 17 for the Low and High Single-Test Groups (*N* = 25 in each group) and in Table 18 for the All-Tests Group (*N* = 50).

The first thing to notice in Table 17 is that the intercorrelations are on the whole much higher in the Day 1 data than in Day 2, especially for the Low Group. It could be assumed that this increase in intercorrelations is due to the increasing stability of *Ss'* performance with increasing practice. But then look at the intercorrelations in the All-Tests Group (Table 18). This Group has had a good deal of practice in serial learning of color-form and trigram lists, since the Serial Intralist Interference test, consisting of three color-form and three trigram lists, was administered before the Serial Retroactive Inhibition test. Despite this prior practice, the All-Tests Group show very poor correlations among the variables in Table 18. Almost the only substantial correlations are those between the error and oscillation scores. It should be noted, incidentally, that the correlation between errors and oscillation is

Table 17

Intercorrelations Among Variables in the IL-High and IL-Low Groups ($\underline{N}$ = 25 in each group) for the Serial Retroactive Inhibition Paradigm. (Day 1 above the diagonal, Day 2 below.)

			IL-Low Group			
Measures	1	2	3	4	5	6
1. OL-Errors		.43	.41	-.33	.39	.21
2. IL-Errors	.24		.28	.25	.37	.14
3. RL-Errors	.14	.60		-.05	.41	.47
4. OL-Oscillation	.03	.23	.39		-.12	-.10
5. IL-Oscillation	.45	.61	.49	.36		.13
6. RL-Oscillation	.41	.55	.79	.36	.59	

			IL-High Group			
Measures	1	2	3	4	5	6
1. OL-Errors		.43	.37	.14	.11	.25
2. IL-Errors	.53		.46	-.11	.31	.44
3. RL-Errors	.74	.64		.02	.12	.30
4. OL-Oscillation	.15	.29	.08		-.39	.19
5. IL-Oscillation	.43	.35	.46	.55		.38
6. RL-Oscillation	.36	.36	.47	.47	.69	

Table 18

Intercorrelations Among Variables in the All-Tests Group ($\underline{N}$ = 50) for the Serial Retroactive Inhibition Paradigm

(Decimals Omitted)

Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. OL-L[1] (Errors)		47	20	40	25	05	47	14	06	32	28	16
2. OL-H[1] (Errors)			12	35	00	16	05	38	08	12	-01	16
3. IL-L (Errors)				22	31	11	16	16	74	20	24	04
4. IL-H (Errors)					14	41	22	19	14	51	-05	04
5. RL-L (Errors)						04	11	08	24	24	84	10
6. RL-H (Errors)							07	11	03	38	-08	55
7. OL-L (Oscillation)								18	10	47	14	04
8. OL-H (Oscillation)									08	35	-01	-33
9. IL-L (Oscillation)										13	16	-08
10. IL-H (Oscillation)											12	26
11. RL-L (Oscillation)												29
12. RL-H (Oscillation)												

[1] "L" refers to the Low Interference condition (i.e., IL is the trigram list); "H" refers to the High Interference condition (i.e., IL is a color-form list).

much lower in the Single-Test than in the All-Tests Group. Also, the magnitude of these correlations in both Groups indicates that errors and oscillation are far from being redundant measures.

Since a correlation of at least .27 is necessary for significance at the .05 level when *N* = 50, it can be said that IDs have little variance in common from one list to another in this test, at least in the All-Tests Group. It was expected that, because of the greater amount of practice in serial learning had by the All-Tests Group, their correlations would be higher than those for Day 2 of the Single-Test Group. But quite the contrary occurred. The average level of performance of the All-Tests Group was not very different from Day 2 of the Single-Test Groups, nor was the variance very different in the two groups (see Table 15). The meager intercorrelations in Table 18 thus seem theoretically unaccountable. Further discussion of this problem is postponed until the section dealing with the factor analysis of all the tests.

C. The Stroop Color-Word Test

Scoring. - A part of this study consisted of a comprehensive review of the literature involving the Stroop Test. Since inclusion of this review in this Report would add some 100 pages to its length, the review is being published separately (Jensen & Rohwer, in press).

Fourteen different Stroop scores were found in the literature; three of them are the basic time scores on each of the three cards; the other eleven are based on a variety of scoring formulas into which are entered the basic time scores. All kinds of claims have been made for the various scoring formulas and the question immediately arose which scores were most appropriate for the purposes of the present study. It was obvious that there was a great deal of redundancy among these derived scores, and the first task consisted of reducing this redundancy. To do this our Stroop data were scored by all the formulas and these scores were intercorrelated and factor analyzed. Table 19 shows the scores that were entered into the factor analysis.

The factor analysis was based on a total of 436 *Ss*, each of whom took the Stroop test on two occasions. The analysis was performed on the first and second administrations separately, and since they showed no appreciable differences for any of the scores, the two administrations were combined and the factor analysis was performed on the composite scores. The same analysis was carried out on the 50 *Ss* in the All-Tests Group, whose Stroop scores were a composite of ten administrations. The results of this analysis were practically identical to those for the Single-Test Groups, so only the latter are reported here.

The correlations between the scores were subjected to a principal components analysis, which was rotated to simple structure by the varimax method. The results are shown in Table 20. Only three factors account for all the variance in the various Stroop scores. The varimax solution is particularly clear-cut and makes for easy interpretation of the factors.

Factor I, which is most unambiguously represented by Score D (see Table 19) can best be regarded as a color-naming factor. Score D is a measure of IDs in the degree of difficulty *Ss* have in naming colors, with the speed factor (i.e., speed of reading words) partialed out.

Table 19

Basic Stroop Scores and Scoring Formulas

Basic Scores	Time Measures (Seconds)
A	Color Card
B	Word Card
C	Color-Word Card

Derived Scores	Scoring Formulas
D	$A/(A + B)$
E	$C - A$
F	A/B
G	A/C
H	$A - B$
I	$(A - B)/(A + B)$
J	B/A
K	$(C - A)/B$
L	$C_{\underline{z}} - 2A_{\underline{z}} + 10$[1]
M	$B(C - A)/A$
N	$B(C - A)/AC$

[1] The C and A raw scores are converted to *z* scores in this formula.

Factor II, represented most clearly by Score E, is the interference factor. It is a measure of the increment in difficulty of color naming brought about by the interference of the conflicting printed words. It can be regarded as a practically pure measure of interference due to response competition, the stronger habit (word reading) blocking the weaker (color-naming). The factor analysis shows that whatever it is that causes color naming per se to be more difficult than word reading, it is not at all the same kind of difficulty that makes for slowness on the color-word interference card.

Table 20

Principal Components and Varimax Rotation of Stroop Scores (First + Second Administration) (*N* = 436)

	Principal Components			Varimax Solution		
Score	I	II	III	I	II	III
A	73	37	57	57	01	82
B	-08	19	97	-34	06	93
C	29	88	37	38	66	64
D	93	23	-29	100	-03	03
E	-20	97	03	07	97	24
F	92	24	-28	99	-04	04
G	59	-73	28	24	-94	23
H	94	32	03	93	-02	36
I	93	24	-29	100	-03	03
J	-92	-23	30	-99	05	-02
K	-16	88	-42	23	94	-22
L	-79	09	-56	-50	43	-71
M	-58	79	18	-37	90	21
N	-90	42	-04	-68	69	-17
% Var.	51	31	17	46	33	20

Factor III is the only factor which is most clearly represented by one of the basic scores - Score B. It is best regarded as a reading speed factor.

It was decided to use only scores B (speed factor), D (color factor) and E (interference factor) in all subsequent analyses. These three scores contain all the essential information that can be derived from the Stroop Test. These scores have the highest loadings on each of their respective uncorrelated factors and are only very slightly intercorrelated with each other.

In all subsequent correlational analyses these three scores were reflected, i.e., their sign was changed, so that the speed factor (symbolized as Sp) means that *Ss* with high scores are superior (i.e., fast) in reading the word card. The interference factor (Int) means *resistance* to interference. The color-speed factor (C-Sp) means that *Ss* with high scores are fast at color-naming (with word speed factor partialed out).

e. *Effects of Practice.* - There are small but highly reliable practice effects on the Stroop for all scores. The effects of practice are negligible after the third administration, and even between the first and second administration there is little interaction between *Ss* and practice effects, so that the test-retest reliability is quite good for all scores. The subject of practice effects on the Stroop has been discussed at length elsewhere (Jensen, in press; Jensen & Rohwer, in press). Table 21 shows the analysis of variance of the three basic time scores obtained in ten administrations to the All-Tests Group. It can be seen that the one effect that is not at all significant is the *Ss* X Administrations interaction. The tremendous *F* ratio for the Scores main effect merely reflects the fact that not one subject in the entire sample of 400+ had scores that departed from this order of difficulty: color-words most difficult, color naming intermediate, word reading least difficult.

Table 21

Analysis of Variance of Basic Stroop Stroop Scores (A, B, C) for 10 Administrations

Source of Variation	MS	*df*	*F*	% Variance
Scores (S)	213,230.38	2	874.14**	43.92
Administrations (A)	5,228.50	9	20.69**	4.85
Subjects (*Ss*)	1,947.44	49	7.98**	9.83
S x A	1,441.80	18	5.91**	2.67
S x *Ss*	481.95	98	1.93*	4.86
A x *Ss*	232.76	441	< 1	11.48
Residual	243.93	882		22.16

* *p* < .01
** *p* < .001

b. Descriptive Statistics of Ss' Performance. - The means and SDs of the three basic and two derived scores are shown below, based on *N* = 436.

Score	First Administration Mean	First Administration SD	Second Administration Mean	Second Administration SD
A	58.24 sec.	10.17	56.11 sec.	10.57
B (Sp)	38.09 "	5.84	37.19 "	5.62
C	100.36 "	19.50	88.34 "	15.93
D (C-Sp)	.60	.04	.60	.04
E	42.12	14.96	33.23	10.50

c. Reliability of the Scores. - The reliabilities of the Stroop scores for the Single-Test Groups (*N* - 436) are based on the correlation between the two administrations of the test, boosted by the Spearman-Brown formula to give the reliability of the composite score based on two administrations. (All subsequent correlational analyses are based on the composite scores.) The reliabilities of the All-Tests Group (*N* = 50) are for the composite of ten administrations. These composite scores were used in the factor analysis of the All-Tests data.

Stroop Score	Reliability Two Administrations	Reliability Ten Administrations
A. Color Card	.89	.98
B. Word Card (Sp factor)	.94	.98
C. Color-Word Card	.83	.98
D. A/(A+B) (C-Sp factor)	.83	.97
E. C-A (Int factor)	.65	.93

d. Correlations of Stroop Scores with Other Variables. - Table 21 shows the correlations of the three Stroop factors with all the memory span tests used in this study. The results are puzzling. The Immediate and Delayed Digit Span and the Immediate and Delayed Color-form Span Tests show quite significant correlations with the Stroop scores in the Single-Tests Groups, and these are "cross-validated" in the All-Tests Group. Beyond this, however, the results are ambiguous. Why the All-Tests Group shows significant correlations between these measures, is a mystery. The number and magnitude of the significant correlations, however, can leave little doubt that there is some relationship between the Stroop and memory span. But the nature of this relationship is is entirely unclear from these data. It is evident, however, that the Interference factor of the Stroop does not coincide with the kind of interference that has operated in the memory span tasks. There is little or no relationship between Stroop Interference and the degree of interference we have hypothesized as operating in the memory span tasks. For example, the retroactive inhibition of memory span (Test Nos. 9 and 10 in Table 22) shows no appreciable correlation with the Interference factor.

Table 22

Correlations of Stroop Factors with Memory Span Tests (Position Scores) in the Single-Test and All-Tests Groups (Decimals omitted)

			Single-Test Group			All-Tests Group (N = 50)		
Test	Day	N^1	Sp^2	Int^3	$C\text{-}Sp^4$	Sp	Int	C-Sp
1. IDS, 2-9	1	50	56**	34*	-37**	39**	21	-22
2. "	2	"	40**	28*	-23			
3. DDS, 2-9	1	"	64**	42**	-46**	40**	28*	-22
4. "	2	"	53**	37**	-32*			
5. ICFS, 2-9	1	30	27	39*	23	26	14	-36**
6. "	2	"	43*	40*	-01			
7. DCFS, 2-9	1	"	31	35*	38*	22	-01	-25
8. "	2	"	34	38*	22			
9. DS-RI, 4-7	1	52	05	18	17	31*	16	-13
10. "	2	"	06	17	19			
11. DS-PI, 4-7	1	"	13	-03	07	31*	24	-06
12. "	2	"	07	-16	-08			
13. DDS, 4-7	1	"	01	-08	-03	32*	13	-03
14. "	2	"	05	-17	-03			
15. IDS, 4-7	1	"	-02	-04	00	24	05	-07
16. "	2	"	02	-15	-02			
17. FDS-Pre	1	30	-17	08	06	27*	04	-17
18. "	2	"	-09	09	15			
19. FDS-Post	1	"	-16	08	08	09	12	08
20. "	2	"	-08	04	17			
21. RDS-Pre	1	"	-02	-03	09	22	03	-03
22. "	2	"	27	01	-08			
23. RDS-Post	1	"	-15	09	08	18	08	-14
24. "	2	"	-01	07	04			
25. DS, 5-12	1	30	07	28	15	24	07	00
26. "	2	"	08	47**	27			
27. RDS, 5-12	1	"	-02	21	15	34*	07	-11
28. "	2	"	-04	39*	32			

[1] Dittos indicate that the particular test was taken by the same group of _S_s that took the preceding test.

[2] Speed factor. A positive _r_ means that good performance on the memory span task is associated with speed of reading words on the Stroop.

[3] Resistance to interference factor. A positive _r_ indicates that good memory span performance is associated with low susceptibility to Stroop interference.

[4] Color-naming speed factor. A positive _r_ indicates that good memory span performance is associated with speed of naming colors (with word reading partialed out).

* $P < .05$
** $P < .01$

Table 23 shows the correlations between the Stroop factors and the serial learning tests. These, too, are puzzling. For example, why should the Serial Retroactive Inhibition Tests not show correlations with the Stroop while the Serial Intralist Interference Tests show a number of significant correlations? Both sets of tests consist of color-form and trigram lists. Again, there is the possibility that the factorial composition of these serial learning tasks changes with practice, so that at one stage of practice they correlate with the Stroop and at another stage they do not.

The Stroop scores have no significant correlations with age or sex in the All-Tests Group.

D. Factor Analysis of All Tests

Supplementary Variables. - Table 24 presents the means, *SD*s, and test-retest reliabilities (based on the Single-Test Groups) of the 43 variables that entered into the factor analysis of the results from the All-Tests Group. It will be noted that a few variables have been added that were not previously considered, viz., chronological age of *S*s, the Eysenck Personality Inventory, the college grade point average, and Raven's Progressive Matrices Test.

Age was included because it had been found in previous studies of serial learning that age, even within the very restricted range of the college population, shows a negative correlation with speed of learning. Also, research (reviewed in Section III of this report) has suggested that the negative correlation between age and performance on learning tasks might be due to increased susceptibility to interference in older *S*s. Therefore it seems proper to include age as a variable in our factor analysis even though it is not apt to be a very important source of variance in our college-age sample.

The Eysenck Personality Inventory, which has equivalent Forms A and B, was administered to supplement the Maudsley Personality Inventory. These Inventories all measure Neuroticism (N) and Extraversion (E). The EPI is an improvement over the MPI largely because it achieves complete orthogonality of the two factors, N and E, which are slightly correlated in the MPI. With three scales each for measuring N and E, these personality factors should be solidly represented in the factor analysis, and the contribution of these factors to the variability in our performance tests should be clearly discernible. It might be asked why, of all the possible personality measures we could have chosen to include in this analysis, we chose neuroticism and extraversion. There are two main reasons. In the first place, both neuroticism and extraversion are superfactors in the personality domain, that is, they are large, independent dimensions which account for most of the variability in personality. This means that if all existing personality questionnaire items were factor analyzed, the two largest second-order factors would be identifiable as neuroticism and extraversion. The MPI and EPI scales are the most highly developed measures of these factors. The second, and more important, reason for including measures of N and E is that both these personality factors are

Table 23

Correlations of Stroop Factors With Serial Learning Tests in the All-Tests Group (N = 50) (Decimals omitted)

Test	Sp[1]	Int[1]	C-Sp[1]
Errors			
CL-L	-14	02	14
CL-H	-18	-03	08
IL-L	-06	01	-03
IL-H	03	07	28*
RL-L	-13	-23	06
RL-H	-04	01	19
Percent RI	03	10	-03
Oscillation			
CL-L	-14	-05	-09
CL-H	-24	-15	06
IL-L	15	07	-19
IL-H	-06	-04	15
RL-L	-12	-19	-04
RL-H	-19	00	02
Errors			
Words	-45**	-23	12
Tri-A	-31*	-27*	06
Tri-B	-21	-35**	-01
Tri-C	-39**	-41**	10
Sum of Tri	-38**	-43**	07
CF-A	-17	-09	-17
CF-B	-19	-07	-01
CF-C	-41**	-25	26
Sum of CF	-31*	-16	-02
Oscillation			
Words	-33*	-23	27
Tri-A	-21	-10	-04
Tri-B	-30*	-24	11
Tri-C	-32*	-10	06
Sum of Tri	-40**	-20	08
CF-A	-39**	-29*	02
CF-B	-08	00	-03
CF-C	-11	-09	26
Sum of CF	-27*	-16	08

[1] See footnotes 1, 2, and 3 or Table 22.

*P < .05

**P < .01

Table 24

Means, SDs, and Test-Retest Reliabilities of the Variables Included in the Factor Analysis of the All-Tests Group ($\underline{N}$ = 50)

	Variable	Mean	SD	r_{tt}
1.	IDS, 2-9	176.68	21.96	.70
2.	DDS, 2-9	152.52	26.74	.79
3.	DS-RI, 4-7	55.70	12.81	.60
4.	DDS, 4-7	77.90	6.91	.59
5.	DS-PI, 4-7	52.38	13.25	.58
6.	IDS, 4-7	79.58	8.16	.37
7.	DCFS, 2-9	84.00	18.15	.69
8.	DCFS, 2-9	65.18	16.81	.87
9.	DS, 5-12	214.36	43.04	.89
10.	RDS, 5-12	239.28	39.39	.83
11.	FDS-Pre	70.72	22.85	.87
12.	FDS-Post	42.76	17.45	.84
13.	BDS-Pre	73.46	19.14	.92
14.	BDS-Post	63.66	13.83	.88
15.	Serial RI, CL-1 (Errors)	54.52	30.00	.42
16.	" " CL-2 "	53.16	30.46	.60
17.	" " IL-1 "	19.06	11.73	.94
18.	" " IL-2 "	31.88	16.90	.56
19.	" " RL-1 "	8.62	9.59	.52
20.	" " RL-2 "	29.88	16.06	.32
21.	" " CL-1 (Oscil.)	7.65	1.17	.20
22.	" " CL-2 "	7.59	1.40	-.09
23.	" " IL-1 "	6.03	1.95	.80
24.	" " IL-2 "	8.06	1.24	.42
25.	" " RL-1 "	3.88	2.88	.34
26.	" " RL-2 "	7.68	1.50	-.05
27.	Serial, Words (Errors)	35.88	31.97	
28.	" Sum of Tri "	109.56	55.54	.71
29.	" Sum of CF "	205.76	90.15	.60
30.	" Words (Oscil.)	7.21	1.87	
31.	" Sum of Tri "	21.93	3.26	.47
32.	" Sum of CF "	24.33	2.67	.56
33.	Age (months)	232.58	15.11	
34.	Stroop, Speed Factor[1]	364.70	52.11	.98
35.	" Interference Factor	253.82	65.27	.93
36.	" Color-Speed Factor	0.58	0.03	.97
37.	MPI-Neuroticism	23.18	7.89	.83
38.	MPI-Extraversion	27.62	8.70	.81
39.	EPI, Form A, Neuroticism	9.00	4.10	.81
40.	EPI, Form B, "	10.16	3.73	.81
41.	EPI, Form A, Extraversion	12.34	4.21	.76
42.	EPI, Form B, "	14.66	3.43	.76
43.	College Grade Point Average	2.41	0.57	
44.	Progressive Matrices (No. correct)	55.16	2.86	.93
45.	Total time (min.) for P.M.	31.16	8.34	

[1] Stroop scores are the total of 10 administrations.

theoretically linked to performance in learning tasks. It should be noted that neuroticism is practically equivalent to anxiety of the type measured by the Taylor Manifest Anxiety Scale (MAS). The N scales of the MPI and EPI correlate almost as highly with the MAS as their reliabilities will permit. Neuroticism has been related to learning through the Yerkes-Dodson Law and through the Spence-Taylor hypothesis, which attributes to anxiety (or neuroticism) the characteristics of drive in Hullian learning theory (Spence, 1964). Extraversion has been related to learning by Eysenck (1957, 1965), in his hypothesis that extraverts build up reactive inhibition (or cortical inhibition) more rapidly and dissipate it more slowly than introverts. Extraverts have been found to differ from introverts in pursuit rotor reminiscence (Eysenck, 1957) and in eyeblink conditioning and extinction (Eysenck, 1965). A previous study by Jensen (1962) found that extreme groups in both extraversion and neuroticism showed differences in serial learning when color-form series were presented at a difficult, rapid rate (2 sec.) but did not differ appreciably when the lists were presented at an easy, slow rate (4 sec.)

The Ss in the present study differ hardly at all from the normative data on the MPI and EPI scales presented in the manuals for these inventories (Eysenck, 1959; Eysenck & Eysenck, 1964). While the college population differs markedly from the general population in intellectual abilities, they do not differ significantly in mean or standard deviation on neuroticism and extraversion.

Raven's Progressive Matrices are quite different in this respect. This intelligence test is practically a pure measure of "g", the general factor common to all intelligence tests. Unfortunately it proved to be a poor test in our present population. Since it was designed and standardized for use with the general population, it probably has relatively little discriminatory power among an intellectually superior sample of the population. It was thought that by having a time limit of 45 minutes the test would be made sufficiently difficult to give some spread of scores even among college students. This was not the case. The Ss completed the test, on the average, in 31 minutes, and their mean score is above the 95th percentile according to the norms provided by Raven (1958). One standard deviation below the mean is the 90th percentile. Obviously a much more difficult test of intelligence should have been used in this study. The Terman Concept Mastery Test is an appropriately difficult test and was administered to over 150 Ss in the Single-Test Groups, but analysis of these data in relation to the performance tests will be published separately. Despite the limitations of the Raven in this study, we shall see that it actually makes a fair showing in the factor analysis.

Grade point average (GPA) was included mainly out of curiosity to see if what we were measuring in the laboratory had any "real-life" correlates. Not much hope was held out for a significant relationship between GPA and the experimental tests, since GPA was treated in a rather crude manner, different Ss' averages being based on different numbers of semesters of work at the University, on different majors, on grade point averages transferred from other colleges, and so on. With only

50 *S*s in the group refinements of the GPA would have been impossible. The reliability of GPA has generally been found to be in the region of .65 to .75. Grade point average is figured on the following basis: A=4, B=3, C=2, D=1. The mean of 2.41 indicates a not quite C+ grade average for the All-Tests Group.

1. Principal Components Analysis.

All the factor analyses were carried out on both Position Scores and Sequential Scores. No essential differences between the two types of scores showed up in the factor analyses; in fact, the two sets of results are practically identical. Therefore only the results based on Position scores are presented here.

The matrix of zero order correlations (Table 25) was first subjected to a principal components analysis (Table 26). Twelve significant components were extracted, the last having an Eigenvalue of 1.26. Eigenvalues of 1.00 or greater are generally considered well within the region of statistical significance. The first twelve principal components of this correlation matrix (Table 25) account for 77 percent of the total variance; the percent contribution of each component is shown in the last row of Table 26. A principal components analysis rarely lends itself to detailed psychological interpretation before rotation and serves mainly as a means of arriving at a psychologically meaningful rotated solution. Before leaving the principal components analysis shown in Table 26, however, a number of points are worth noting:

a. The communalities (h^2), or proportion of common factor variance in each test, are generally very high, indicating a healthy condition for factor analysis. The communality of a test may be regarded as the lower bound of the test's reliability, for the reliability of a test cannot be lower than its communality. Yet we see that a number of the serial learning tests which had quite mediocre test-retest reliabilities have quite high communalities. This can only mean that although performance on equivalent forms of a serial learning task does not show high stability, the performance on each task is itself quite reliable. Thus it appears very likely that the factorial composition of serial learning changes with practice. The intercorrelations between equivalent forms of serial learning tasks will therefore be rather low, even though the internal consistency reliability of each task, as well as their communalities, may be quite high. It is also interesting that the changes in factorial composition of serial learning tasks as a function of practice seems to take place between successive tasks and not during the course of learning any one serial list. These indications in the present data point up problems in the study of IDs in serial learning (and probably other types of verbal learning) that will have to be investigated in their own right. A comprehensive account of IDs in learning cannot consider only the genotypic sources of IDs in first-task learning, for the chief sources of variance in the first learning session (even with highly "equivalent" forms) may not be the same as the sources of ID variance in the second, third, or subsequent performances on the task. Because of this lack of factorial stability over the course of practicing on several serial lists, it will be seen in the present study that, even though a good deal of the serial learning variance is accounted for among the twelve principal components (as indicated by the high communalities), this variance is so fragmented and scattered among the various components (and among the factors after

Table 25

Correlations Among Variables in the All-Course Group (N = 30) (Decimals Omitted)

Variable	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. IDS,2-9	82	43	56	40	43	43	46	52	54	47	31	44	36	-05	-08	-15	-08	-10	01	-05	-01	-04	-13	-18	04	-23	-31	-24
2. DDS,2-9		53	48	38	33	44	57	56	59	56	34	26	24	-23	-05	-18	-30	-10	-05	-10	-05	-10	-19	-25	14	-40	-45	-17
3. DS-RI,4-7			48	28	37	50	58	59	61	74	46	30	13	-17	-14	-11	-16	18	-05	-11	-01	05	-08	07	04	-36	-25	-17
4. DDS,4-7				53	65	31	38	47	54	43	10	25	23	-04	-13	-16	-09	-05	-17	02	02	-12	-07	-08	-08	-35	-16	-19
5. DS-PI,4-7					55	26	34	44	45	36	37	38	34	-06	-01	-17	-30	-07	-07	-08	08	-13	-06	-13	12	-24	-08	-23
6. IDS,4-7						19	15	45	46	22	10	32	35	-08	-22	-16	-06	-04	-13	00	12	-08	-08	-05	-14	-35	-13	-13
7. ICPS,2-9							81	46	49	51	07	20	27	-04	-17	-05	-31	-11	-16	13	-25	-05	-10	-05	06	-20	-12	-13
8. DCPS,2-9								56	52	66	36	24	23	-13	-02	-05	-02	01	-06	04	-08	02	-11	-02	10	-20	-10	-06
9. DS,5-12									89	72	53	38	48	-26	-24	-21	-30	-08	01	-17	-20	-14	-20	-13	10	-32	-38	-38
10. RDS,5-12										76	56	40	48	-24	-26	-21	-34	-10	02	-17	-19	-13	-17	-13	04	-44	-41	-40
11. FDS-Pre											50	32	31	-06	-07	-10	-02	-02	-07	-06	-06	07	-14	-04	08	-26	-21	-19
12. FDS-Post												24	30	-05	-02	-04	-13	-10	10	-13	01	-16	-01	-04	13	-37	-22	-05
13. BDS-Pre													54	-12	-13	-32	-41	-14	-11	-18	-13	-24	-18	-01	-05	-31	-27	-47
14. BDS-Post														-26	-38	-27	-36	-23	-02	-21	-17	-13	-13	-10	00	-20	-34	-44
15. Ser. RI, CL-1 (Errors)															47	20	41	24	-05	47	14	05	32	28	-16	27	34	42
16. " " CL-2 "																12	57	00	-16	05	38	-08	12	-01	-16	21	20	61
17. " " IL-1 "																	62	31	12	16	16	74	20	24	-04	22	43	37
18. " " IL-2 "																		16	41	22	19	14	51	-05	04	17	10	23
19. " " RL-1 "																			04	11	08	24	24	04	-10	-02	43	18
20. " " RL-2 "																				07	11	07	38	-08	77	-04	13	-03
21. " " CL-1 (Oscil.)																					18	10	47	14	-04	28	20	18
22. " " CL-2 "																						08	20	-01	-12	12	27	30
23. " " IL-1 "																							13	16	-08	12	27	11
24. " " IL-2 "																								12	26	07	34	23
25. " " RL-1 "																									-23	-12	43	10
26. " " RL-2 "																										09	07	-08
27. Ser. Words (Errors)																											30	10
28. " Sum of Tri "																												45
29. " Sum of CF "																												
30. " Words (Oscil.)																												
31. " Sum of Tri "																												
32. " Sum of CF "																												
33. Age (months)																												
34. Stroop, Speed Factor																												
35. " Interference Factor																												
36. " Color-Speed Factor																												
37. MPI-Neuroticism																												
38. MPI-Extraversion																												
39. EPI, Form A, Neuroticism																												
40. EPI, Form B, "																												
41. EPI, Form A, Extraversion																												
42. EPI, Form B, "																												
43. College Grade Point Average																												
44. Progressive Matrices																												
45. Total time for P.M.																												

Table 25 Continued

Variable	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1. IDE,2-9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2. DDE,2-9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3. DE-RT,4-7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4. DDE,4-7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5. DE-FT,4-7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6. IDE,4-7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7. ICPS,2-9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8. DCPS,2-9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9. DE,5-12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10. RDE,5-12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11. PDE-Pre	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12. PDE-Post	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13. RDE-Pre	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14. RDE-Post	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15. Sep. RE, CL-L (Errors)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16. " " CL-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17. " " II-L "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18. " " II-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19. " " RL-L "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20. " " RL-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21. " " CL-L (Deptl.)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22. " " CL-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
23. " " II-L "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24. " " II-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25. " " RL-L "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
26. " " RL-H "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
27. Sep. Words (Errors)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
28. " Sum of Tri "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
29. " Sum of CP "	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
30. " Words (Deptl.)		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
31. " Sum of Tri "			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
32. " Sum of CP "				[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
33. Age (months)					[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
34. Stroop, Speed Factor						[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
35. " Interference Factor							[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
36. " Color-Speed Factor								[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
37. MPI-Neuroticism									[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
38. MPI-Extraversion										[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
39. EPI, Form A, Neuroticism											[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
40. EPI, Form B, "												[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
41. EPI, Form A, Extraversion													[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
42. EPI, Form B, "														[illegible]	[illegible]	[illegible]	[illegible]
43. College Grade Point Average															[illegible]	[illegible]	[illegible]
44. Progressive Matrices																[illegible]	[illegible]
45. Total time for P.M.																	

Table 26

Principal Components and Communalities (h^2) of the Variables in the All-Tests Group
(Decimals omitted)

Variable		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	h^2
1. IDS,2-9	(01)	66	08	51	08	00	04	06	-12	-05	-05	15	-14	76
2. DDS,2-9	(02)	70	03	45	03	04	20	04	03	-03	-14	17	16	81
3. DS-RI,4-7	(03)	60	44	20	-03	-03	12	-08	12	-06	04	24	15	71
4. DDS,4-7	(04)	73	22	26	13	00	-32	23	-15	12	20	01	17	77
5. DS-PI,4-7	(05)	36	17	23	07	10	-32	16	-14	00	-04	-32	10	67
6. IDS,4-7	(06)	50	07	26	13	-07	-47	22	-38	-02	-01	-07	08	77
7. IDFS,2-9	(07)	54	38	16	-02	-16	30	08	-01	28	06	-06	-12	74
8. DDFS,2-9	(08)	55	45	26	-10	-10	32	07	10	20	-09	16	08	73
9. DS,5-12	(09)	73	23	04	-20	11	-08	-02	04	-06	-01	03	01	73
10. RDS,5-12	(10)	64	29	03	-10	10	-08	02	02	-01	01	11	-04	82
11. FDS-Pre	(11)	66	30	11	-07	03	13	10	13	02	04	20	02	82
12. FDS-Post	(12)	51	38	-01	-04	23	-05	-03	43	-22	-13	-02	-13	72
13. BDS-Pre	(13)	77	-06	03	-20	-22	-02	-06	03	-07	13	-06	-30	76
14. BDS-Post	(14)	60	-02	-12	-15	-07	-17	-14	-17	06	-10	-17	-45	73
15. Ser. RI, CL-1 (Errors)	(15)*	43	-26	-42	-15	11	14	02	-20	-10	-06	23	08	70
16. " " CL-2 "	(16)*	36	-04	-41	-09	10	01	-28	-54	23	20	13	04	80
17. " " IL-1 "	(17)*	41	-47	-28	-38	-06	-18	08	22	17	00	-16	15	80
18. " " IL-2 "	(18)*	43	16	-44	-22	-41	-10	26	-01	07	-16	-16	07	81
19. " " RL-1 "	(19)*	24	-43	-08	-15	30	33	36	01	-07	04	-20	-01	82
20. " " RL-2 "	(20)*	11	-06	-02	07	-73	-01	38	05	-05	13	-14	17	75
21. " " CL-1 (Oscil.)	(21)*	31	-30	-33	-07	-03	-06	-04	08	-46	-23	31	14	69
22. " " CL-2 "	(22)*	30	-17	-41	-04	-04	23	-32	03	26	13	-02	11	61
23. " " IL-1 "	(23)*	43	-37	08	-37	02	-05	10	44	33	-16	-20	03	82
24. " " IL-2 "	(24)*	36	-03	-33	-16	-41	07	30	05	-04	-01	13	18	76
25. " " RL-1 "	(25)*	20	-37	03	-24	48	33	30	-14	-14	00	-13	-04	53
26. " " RL-2 "	(26)*	-01	-16	-08	46	-35	-05	18	21	-03	33	-05	-10	81
27. Ser. Words (Errors)	(27)*	61	-09	-01	38	06	-17	-24	12	-08	-10	01	02	67
28. Ser. Sum of Tri "	(28)*	64	-52	-06	-05	07	19	-03	04	-16	05	00	15	78
29. " Sum of CF "	(29)*	53	-24	-03	-03	05	-05	-18	-33	07	35	06	-05	73
30. " Words (Oscil.)	(30)*	54	-05	-04	13	-31	05	-48	31	07	-17	-29	-09	86
31. " Sum of Tri "	(31)*	55	-17	-03	-20	-06	30	-30	26	04	08	-18	10	77
32. " Sum of CF "	(32)*	57	-20	-36	01	-16	-10	-10	00	-41	04	13	-10	76
33. Age (months)	(33)	-38	26	-11	03	-16	48	13	13	-13	30	10	-42	73
34. Stroop, Speed Factor	(34)*	53	-11	11	44	-02	29	-18	-15	-14	23	-18	10	73
35. " Interference Factor	(35)*	28	-17	12	30	18	30	-10	02	-20	29	-34	29	71
36. " Color-Speed Factor	(36)*	-22	-08	02	-31	37	-44	-07	26	-15	35	11	10	74
37. MPI-Neuroticism	(37)	28	05	-54	57	30	-06	22	14	08	-01	-07	-16	90
38. MPI-Extraversion	(38)	17	-74	33	13	-03	00	-06	04	16	06	23	-11	87
39. EPI, Form A, Neuroticism	(39)	33	24	-41	38	30	00	18	13	16	04	-05	-02	86
40. EPI, Form B, "	(40)	26	19	-55	49	19	-17	05	02	33	01	02	13	83
41. EPI, Form A, Extraversion	(41)	16	-74	15	24	-07	07	14	02	26	05	29	04	83
42. EPI, Form B, "	(42)	15	-63	36	29	-16	02	00	10	28	04	21	-05	81
43. College Grade Point Average	(43)	18	06	30	-19	16	-16	-26	-04	-17	45	-17	15	57
44. Progressive Matrices	(44)	02	08	-31	-25	19	-20	25	30	16	41	36	02	72
45. Total time for P.M.	(45)*	19	-45	19	27	29	-19	-04	06	-21	-13	09	-21	59
Percent of Total Variance		21	10	08	06	05	05	04	04	04	03	03	03	

* The sign has been reversed on all variables marked with an asterisk. All the factor loadings of the memory span, serial learning, and Stroop tasks represent correlations with "goodness" of performance.

rotation) that no very clear picture of the factorial structure of serial learning can emerge. Another study will eventually have to be carried out in which 15 or 20 or more "equivalent" serial lists will be learned by a group of *Ss*; and the intercorrelations among these would have to be factor analyzed to give some indication of the number of independent factors that are involved in serial learning throughout the course of prolonged practice. Identification of these factors would, of course, depend upon including a number of well-chosen "reference" tests in the analysis. The present study suggests what some of those reference tests might be.

The communality of Grade Point Average (GPA) is .57, which is the lowest communality in the lot. A communality may be regarded also as the lower bound of the square of the multiple correlation (*R*) between a particular variable and all the other variables in the matrix. This means that the *R* between GPA and the 44 other variables in this matrix could not be less than √.57 or .76, which is remarkable, considering that the reliability of the GPA is probably no higher than this. The standard error for this correlation, with *N* = 50, is something less than .14. It might be asked whether the one standard test of intelligence included in this battery, viz. Raven's Progressive Matrices, is largely responsible for this remarkable lower-bound multiple-R of .76 between GPA and all the other tests. The answer is that the correlation between the Progressive Matrices and GPA is -.01. The power to predict GPA obviously comes from other tests in the battery. This finding should not be misconstrued as a criticism of the Progressive Matrices Test. As was pointed out previously, the difficulty level of the Progressive Matrices is not appropriate to the present sample of *Ss* and much of the small variance in Matrices scores is contributed by non-intellectual factors.

b. The first principal component can be regarded as the general factor of the matrix. It is interesting to note the loadings of the various tests on this general factor. First of all, we see that the memory span tasks are generally the most heavily loaded on this "g" factor. But many of the serial learning tasks have comparable loadings, and this is true whether the serial learning consisted of words, trigrams, or color-forms. The only serial tasks with quite negligible loadings are those involving relearning of a serial list, especially when the relearning takes place after a highly interfering interpolated list. Though there is a great deal of true variance in RL scores (in the high retroactive interference condition), practically none of this variance is associated with the general factor. Different factors are involved in the relearning of a retroactively inhibited list than are involved in the original learning of the list. This is an important finding worthy of further investigation in future experiments.

The Stroop speed factor has a substantial loading (.53) on the general factor. (This is corroborated by the significant correlations between some of the memory span tasks and the Stroop speed score in the Single-Test Group as well.) How can we account for this relationship between two such phenotypically different tests as speed of word reading on the Stroop and performance on memory span and serial learning tasks? An explanation in terms of general motivational level is one possibility, but nothing is known about the effects of drive on Stroop speed and there is no evidence that drive affects memory span. A more plausible hypothesis is that speed of word reading may involve the same processes as rehearsal or "recirculation" in short-term memory tasks. The faster the *S*'s speed of repeating material, overtly or covertly, the more rehearsal it will receive and the better it will be retained. Or the speed factor may reflect IDs in speed of processing sensory input. The faster that sensory input can be processes to issue in some form of response, the more time there is for rehearsal and consolidation of the trace. The Stroop speed factor is worth further investigation in this connection.

No very enlightening label can be attached to the general factor; it represents short-term learning ability, which in this battery of tests probably largely reflects resistance to interference effects. The Repeated Digit Span Test (5-12) has the highest loading (.84) on the general factor. It was originally hypothesized that this test would be loaded on a short-term learning factor, and since the digit series are of supra-span length on this test, high scores must also reflect resistance to interference with the stimulus trace.

Finally, note that age is negatively loaded (-.32) on the general factor.

c. Significance of factor loadings. - Before the factor analysis was completed it was thought that it might be desirable to correct the factor loadings for attenuation, which is accomplished by dividing the loading by the square root of the reliability of the variable. The final rotated oblique factors turned out so clearly, however, that the picture would not have been appreciably improved by correcting for attenuation, and the interpretation of the factors would not be changed in the least by this correction. Appendix E shows the factor loadings corrected for attenuation. It can be seen that when the reliability of the test is above .70 there is not much effect of correcting for attenuation.

The statistical significance of factor loadings is more problematic, but there are various ways of obtaining a rough and ready estimate of the statistical significance of a factor loading. The method suggested by Cattell (1952, p. 293) for calculating the standard error of a loading (after correction for attenuation) is adopted here:

$$\text{S. E.} = \frac{(1 - r^2)\,ah}{2r\ N}$$

a = factor loading of the variable
r = reliability of the variable
N = number of *S*s

The standard errors of factor loadings corrected for attenuation are presented in Appendix F.

2. Rotated Oblique First-Order Factors.

Both orthogonal and oblique rotations to approximate simple structure were performed on the principal components. Since from a psychological standpoint it is reasonable to expect that factors in the learning domain are likely to be correlated to some extent with one another, it seemed desirable to allow for the emergence of oblique (correlated) factors. A closer approximation to simple structure is also possible when the factors are not arbitrarily forced to remain orthogonal (uncorrelated). The Promax method of analytical rotation to oblique simple structure was used. This method has been discussed in detail elsewhere (Hendrickson & White, 1964). The Promax solution to rotation has been shown to correspond very closely to results arrived at by highly expert visual methods of rotation to simple structure.

Since the factor loadings of oblique factors are not equivalent to the Pearson correlation coefficient, we have also computed the Pearson r between every test and every factor. These correlations help to elucidate the factorial composition of the tests in our battery. The interpretation of the factors, however, is based mainly on the oblique factor loadings.

Interpretations of the rotated factors are based upon loadings of .40 or greater. The 12 first-order factors and their interpretations are presented below. The same format will be maintained throughout. The variables are listed in descending order of magnitude of their factor loadings. An asterisk following the factor loading indicates that the variable had its highest loading on this factor. In the last column are shown the correlations of the variables with the factor. It will be noticed that in many cases the correlations are higher than the factor loadings (and vice versa). Factor loadings of less than .40 are always separated from those greater than .40, which provide the main basis for interpreting the factor.

Table 27 presents the factor analysis. Table 28 gives the correlations of the variables with the factors.

Factor A

	Variable	Loading	Correlation
(08)	Delayed Color-form Span	.98*	.83
(11)	Forward Digit Span, Pre-cued	.95*	.87
(03)	Digit Span: Retroactive Inhibition	.90*	.80
(02)	Delayed Digit Span, 2-9	.79*	.76
(07)	Immediate Color-form Span	.70*	.74
(10)	Repeated Digit Span, 5-12	.65*	.82
(09)	Digit Span 5-12	.64*	.80
(01)	Immediate Digit Span 2-9	.60*	.66
(12)	Forward Digit Span, Post cued	.50*	.56
(04)	Delayed Digit Span 4-7	.40	.55
(05)	Digit Span: Proactive Inhibition	.10	.44
(13)	Backward D.S., Pre-cued	.09	.40
(14)	Backward D.S., Post-cued	-.07	.37

Interpretation: Resistance to retroactive interference in short-term memory (STM).

3-3

Table 27

Totated Oblique Factor Coefficients of First-Order Factors (Decimals omitted)

Variable		A	B	C	D	E	F	G	H	I	J	K	L
1. IDS,2-9	(01)	60	31	06	-16	07	-22	06	-18	01	-14	02	-24
2. DDS,2-9	(02)	79	25	15	-15	06	-15	04	-03	-14	-12	-13	18
3. DS-RI,4-7	(03)	90	-04	06	-07	-03	-02	-18	-14	-11	11	-07	14
4. DDS,4-7	(04)	40	08	-03	16	-23	-65	00	02	21	17	-12	04
5. DS-PI,4-7	(05)	10	-22	12	11	-07	-68	14	19	16	-09	-22	-09
6. IDS,4-7	(06)	04	02	-03	02	-18	-79	00	-08	03	-08	02	-21
7. ICFS,2-9	(07)	70	00	-13	04	-06	09	10	02	41	-23	-01	-14
8. DCFS,2-9	(08)	98	04	06	-06	03	08	-04	02	16	-12	15	07
9. DS,5-12	(09)	64	-16	-08	-01	10	-16	-02	05	-17	14	02	-16
10. RDS,5-12	(10)	65	-08	-11	09	08	-18	-01	02	-11	12	05	-21
11. FDS -Pre	(11)	95	-08	08	08	-01	03	-05	-05	-01	15	06	01
12. FDS-Post	(12)	50	-26	37	20	22	13	07	14	-20	15	01	-19
13. BDS-Pre	(13)	09	08	-15	-18	-10	-01	04	10	06	06	19	-91
14. BDS-Post	(14)	-07	-08	-31	00	14	-11	07	13	03	-21	26	-82
15. Ser. RI, OL-L (Errors)	(15)*	06	02	-38	03	15	02	14	-14	-67	-12	20	07
16. " " OL-H "	(16)*	-12	01	-99	11	26	-10	-11	-23	00	01	01	-20
17. " " IL-L "	(17)*	-04	26	-06	-06	-10	-18	12	86	09	18	10	-06
18. " " IL-H "	(18)*	00	-37	-17	04	-45	-29	05	20	-23	-22	23	-14
19. " " RL-L "	(19)*	-14	02	10	16	-05	08	97	20	03	-05	-15	-17
20. " " RL-H "	(20)*	-02	01	24	-13	-94	-16	-01	10	-08	-11	-14	05
21. " " OL-L (Oscil.)	(21)*	-04	-07	12	-07	-08	07	-03	-07	-97	00	06	09
22. " " OL H "	(22)*	01	09	-58	17	05	30	-12	19	03	03	-25	07
23. " " IL-L "	(23)*	-11	13	28	[illegible]	-01	-12	14	100	14	16	23	-13
24. " " IL-H "	(24)*	13	02	00	-09	-59	07	07	02	-69	07	08	11
25. " " RL-L "	(25)*	-03	-04	06	-01	14	-06	95	06	-05	-01	-24	06
26. " " RL-H "	(26)*	-24	16	25	16	-78	-02	-20	-13	-06	12	-18	-24
27. Ser. Words (Errors)	(27)*	13	17	07	29	10	-17	-24	04	-28	-14	-13	-08
28. Ser. Sum of Tri "	(28)*	10	21	-15	-09	-05	02	25	15	-43	02	-36	06
29. " Sum of CF "	(29)*	-08	10	-81	-08	09	-04	06	-07	-02	18	-12	-43
30. " Words (Oscil.)	(30)*	05	01	06	04	-05	14	-33	41	-01	-34	-17	-33
31. " Sum of Tri "	(31)*	17	-10	-24	-08	-01	32	03	42	-14	02	-34	-06
32. " Sum of CF "	(32)*	-06	-05	-16	02	-21	01	-08	-16	-69	10	-04	-36
33. Age (months)	(33)	18	-04	05	02	-18	72	21	-40	18	06	04	-30
34. Stroop, Speed Factor	(34)*	08	10	-22	14	-12	-08	05	-27	05	-25	-76	05
35. " Interference Factor	(35)*	-09	-11	01	08	-12	-06	25	-10	03	-06	-100	32
36. " Color-Speed Factor	(36)*	-15	-07	03	-10	15	-04	-06	19	-05	77	-10	03
37. MPI-Neuroticism	(37)	-14	00	-01	96	04	01	20	-05	-03	02	-01	-05
38. MPI-Extraversion	(38)	-04	91	-06	-11	06	03	00	11	03	02	04	-12
39. EPI, Form A, Neuroticism	(39)	11	-03	00	94	05	-04	09	-06	11	04	-11	13
40. EPI, Form B, "	(40)	-04	01	-25	88	04	-17	-17	05	07	08	03	23
41. EPI, Form A, Extraversion	(41)	-02	91	-11	09	-09	02	06	13	-05	05	09	14
42 EPI, Form B, "	(42)	04	93	01	01	-08	02	-08	17	14	-03	07	-01
43. College Grade Point Average	(43)	01	-18	-20	-31	04	-21	-05	01	22	34	-61	-06
44. Progressive Matrices	(44)	27	12	-17	26	-06	18	00	16	01	81	27	06
45. Total time for P.M.	(45)*	-17	38	19	12	26	-07	12	-07	-25	00	00	-23

* Asterisks indicate that the variable has been reversed in sign. Thus, in effect, all the memory span, serial learning, and Stroop tasks are keyed in the same direction representing "goodness" of performance.

Table 28

Correlations of Variables With Oblique Factors (Decimals omitted)

Variable		A	B	C	D	E	F	G	H	I	J	K	L
1. IDS, 2-9	(01)	66	36	00	00	12	-49	11	05	-23	-21	-38	-51
2. DDS,2-9	(02)	76	31	-06	02	13	-38	11	23	-33	-28	-44	-31
3. DS-RI,4-7	(03)	80	-10	-10	14	-02	-23	-12	10	-20	-06	-29	-29
4. DDS,4-7	(04)	55	10	-11	23	-14	-71	04	07	-08	07	-22	-32
5. DS-PI,4-7	(05)	44	-04	-01	20	05	-69	05	20	-16	-09	-31	-40
6. IDS,4-7	(06)	33	14	-06	11	-07	-83	02	-01	-17	-06	-15	-40
7. ICFS,2-9	(07)	74	-09	-29	16	-19	-04	09	18	10	-38	-17	-26
8. DCFS,2-9	(08)	83	-10	-17	09	-07	-11	02	18	00	-28	-11	-18
9. DS,5-12	(09)	80	-14	-26	21	09	-40	02	34	-42	-01	-32	-51
10. RDS,5-12	(10)	82	-08	-28	31	05	-43	02	31	-42	-04	-33	-56
11. FDS-Pre	(11)	87	-18	-13	27	-05	-20	-01	16	-17	-03	-22	-35
12. FDS-Post	(12)	56	-21	18	36	27	-11	-05	26	-35	08	-33	-49
13. BDS-Pre	(13)	40	12	-16	03	-11	-25	01	28	-28	00	-26	-80
14. BDS-Post	(14)	37	-02	-34	20	04	-29	03	36	-36	-25	-18	-69
15. Ser. RI, OL-L (Errors)	(15)*	22	07	-54	18	04	-10	29	26	-67	-22	00	-05
16. " " OL-H "	(16)*	17	-05	-79	19	03	-16	02	13	-26	-11	-05	-14
17. " " IL-L "	(17)*	18	28	-31	-05	-08	-20	20	79	-31	02	-12	-17
18. " " IL-H "	(18)*	32	-35	-52	22	-53	-21	11	36	-32	-29	19	-15
19. " " RL-L "	(19)*	08	21	-14	09	-06	02	88	23	-22	-04	-11	-11
20. " " RL-H "	(20)*	08	05	-06	-03	-81	-03	01	08	-01	-16	03	-03
21. " " OL-L (Oscil.)	(21)*	08	08	-12	13	01	-04	05	24	-79	-05	-15	-17
22. " " OL-H "	(22)*	19	03	-60	25	-09	27	-05	44	-26	-19	-26	-05
23. " " IL-L "	(23)*	05	15	00	-03	05	-11	13	77	-20	08	-03	-17
24. " " IL-H "	(24)*	22	04	-37	10	-57	03	21	28	-57	-07	02	-07
25. " " RL-L "	(25)*	10	19	-11	-09	16	-13	91	11	-21	02	-12	03
26. " " RL-H "	(26)*	-14	19	23	23	-61	03	-26	-17	-01	11	-14	-28
27. Ser. Words (Errors)	(27)*	38	27	-01	47	19	-33	-27	28	-53	-25	-52	-50
28. Ser.Sum of Tri. "	(28)*	36	41	-35	08	03	-15	31	50	-69	-16	-54	-32
29. " Sum of CF "	(29)*	28	13	-68	07	-02	-20	17	29	-38	-4	-27	-41
30. " Words (Oscil.)	(30)*	38	10	-09	27	-01	04	-43	58	-36	-49	-54	-56
31. " Sum of Tri "	(31)*	43	-02	-46	12	-03	20	03	70	-47	-21	-47	-30
33. Age (months)	(33)	-06	-19	12	-01	-31	65	13	-41	37	09	18	05
34. Stroop, Speed Factor	(34)*	38	30	-19	29	-03	-16	-02	06	-25	-37	-72	-34
35. " Interference Factor	(35)*	15	19	00	13	07	-05	11	07	-18	-13	-73	-13
36. " Color-Speed Factor	(36)*	-26	-07	20	-30	27	-06	-05	00	01	79	-01	01
37. MPI-Neuroticism	(37)	10	-01	-07	92	-03	-02	08	03	-26	-06	-13	-21
38. MPI-Extraversion	(38)	-06	91	04	-15	13	-13	11	16	-19	-07	-29	-18
39. EPI, Form A, Neuroticism	(39)	27	-08	-07	90	-01	-07	-02	01	-14	-07	-17	-14
40. EPI, Form B, "	(40)	14	-11	-27	84	-07	-12	-18	11	-15	-06	-02	-01
41. EPI, Form A, Extraversion	(41)	-06	85	-13	02	-08	-07	22	19	-24	-10	-13	02
42. EPI, Form B, "	(42)	-02	87	05	-04	-03	-09	03	16	-09	-16	-22	-11
43. College Grade Point Average	(43)	15	-04	-01	-25	19	-25	-10	05	04	32	-46	-25
44. Progressive Matrices	(44)	08	-12	-17	17	-15	10	12	08	-01	66	23	06
45. Total time for P.M.	(45)*	-07	53	28	12	38	-25	09	00	-37	03	-33	-35

* Asterisks indicate that the variable has been reversed in sign.

Discussion: First of all, we see that the rank order of the loadings of these memory span tests corresponds quite closely with the degree to which interference has been hypothesized to be involved in the various tests. Why should the pre-cued forward digit span have a much higher loading on this factor than the post-cued series? It was hypothesized that Ss would gain some subjective impression of their interference tolerance and would "turn off" their attention at some critical point in the forward series in order to protect their memory traces from interference by the succeeding items. This would permit Ss to recall more digits (which they do - see Table 24), but it would also accentuate IDs in resistance to interference. Ss with greater resistance to interference would "turn off" later in the series and would have greater recall.

One of the most interesting findings is the fact that retroactively inhibited (RI) digit span has a loading of .90 on this factor, but proactively inhibited digit span (PI) has a loading of only .10 (though it correlated .44 with this factor). The means, standard deviations, and reliabilities of the RI and PI measures are highly similar. Yet it is clear that RI and PI involve quite different sources of ID variance. It is suggested that this factor might be most fruitfully described as resistance to interference with consolidation of the memory trace. In the case of PI, the consolidation of the second, to-be-recalled list is not retroactively interfered with to any marked degree. The fact that PI is as great in magnitude as RI and yet has so little in common with RI factorially must mean that the performance decrement brought about by the PI paradigm must have some other origin than the kind of interference involved in Factor A. It is hypothesized that the essential difference between RI and PI is that RI involves interference with consolidation while PI does not. The hypothesized basis of PI is a weakening of the initial stimulus trace due to the persisting trace of the previous list. Thus, in the PI paradigm there would be weaker registration of the to-be-recalled list and a lesser trace to be consolidated. The end result of RI and PI look much the same, on the average, yet different sources of IDs are involved in these two forms of performance decrement.

Another point of interest is the extremely low loadings of Backward digit span, both Pre and Post-cued, on this factor. This makes sense, since the last items in a long series (15 digits in this test) suffer the least retroactive inhibition; at the time of recall the memory traces of these last items are still fairly strong. Also, for long series, backward digit span is superior to forward (see Table 24), which is consistent with the idea that less interference is involved in backward recall of series that considerably exceed the S's forward memory span.

Factor B

Variable	Loading	Correlation
(42) EPI, Form B, Extraversion	.93*	.87
(41) EPI, Form A, Extraversion	.91*	.85
(38) MPI, Extraversion	.91*	.91
(45) Total time for Matrices	-.38*	-.53

Interpretation: Extraversion.

Discussion: Oblique rotation to approximate simple structure practically guaranteed that this factor would come out quite exclusively as Extraversion, with only small loadings on other variables. If Extraversion is of any importance in the types of performance measured in this battery, it should show up in the second-order factors or other higher-order factors. The one variable most heavily loaded on Extraversion is the time *Ss* spent taking the Progressive Matrices Test. *Ss* were allowed a limit of 45 minutes, and if no time limit had been imposed this loading on Extraversion might well have been considerably higher than it is. This finding is of particular interest in view of the fact that it is predictable from Eysenck's account of the nature of extraversion. In fact, Professor Eysenck was asked by the writer to offer a prediction concerning the relationship between extraversion and time spent on a cognitive test when there is no imposed time limit, and he made precisely the prediction born out by the present data. Extraverts, being more susceptible to boredom from sustained attention, are in a greater hurry to get the test over with and are less apt to conscientiously recheck their answers, and so forth. They therefore turn in their test papers to the examiner sooner than do the more introverted *Ss*. Extraverts probably pay a slight price for this tendency to finish quickly; their scores are slightly but not significantly lower than the scores of introverts (the correlation between E and Matrices score was -.13 in the present sample).

Correlations between the personality measures and the experimental variables in the various Single-Test Groups are presented in Appendix G.

Factor C

Variable	Loading	Correlation
+(16) Serial RI, OL-H (errors)	-.99*	-.79
+(29) Serial, Sum of CF (errors)	-.81*	-.68
+(22) Serial RI, OL-H (oscillation)	-.58*	-.60
+(15) Serial RI, OL-L (errors)	-.38	-.54
+(18) Serial RI, IL-H (errors)	-.17	-.52
+(31) Sum of Trigrams (oscillation)	-.24	-.46

Interpretation: Not clear. Might be described as poor ability in serial learning of lists with high intralist interference.

Discussion: This factor, like all the others involving mainly serial learning, is not very "clean." All of the variables loaded on it but one (31) consist of color-form series. Relearning of color-form series is not represented on this factor at all. The fact that trigram series correlate -.46 with this factor means it is not specific to

---

[+] Variables marked with a plus (+) sign have been reversed in sign, so that the variable will represent "goodness" of performance. In all subsequent analyses all experimental variables are "keyed" consistently in the direction of "goodness" of performance on the particular task.

color-forms. But only oscillation in the trigram lists is correlated with this factor, which suggests that the factor probably reflects intralist interference tendencies. Interlist or associative interference tendencies are not prominent because interpolated learning (of either color-forms or trigrams) are not appreciably loaded on this factor. Psychologically nothing much can be said about this factor.

Factor D

	Variable	Loading	Correlation
(37)	MPI-Neuroticism	.96*	.92
(39)	EPI, Form A, Neuroticism	.94*	.90
(40)	EPI, Form B, Neuroticism	.88*	.84
+(27)	Serial, Words (errors)	.29*	.47

Interpretation: Neuroticism.

Discussion: As in the case of extraversion, this method of factor analysis almost insured a "pure" Neuroticism factor among the first order factors, with only small loadings of other variables on the factor. The relationships of neuroticism to other factors should show up more clearly in the higher order factors. No specific prediction can be made about the loadings of the present tasks on Neuroticism, for this depends in theory upon the complexity of the task and the amount of response competition it involves. Relatively easy tasks should have positive correlations on Neuroticism and difficult tasks should have negative correlations, while intermediate tasks should have correlations near zero. The present data do not show any clear-cut trends along these lines.

The correlations of Neuroticism with the memory span tasks in the various Single-Test Groups are shown in Appendix G.

Factor E

	Variable	Loading	Correlation
+(20)	Serial RI, RL-H (errors)	-.94*	-.81
+(26)	Serial RI, RL-H (oscillation)	-.78*	-.61
+(24)	Serial RI, IL-H (oscillation)	-.59	-.57
+(18)	Serial RI, IL-H (errors)	-.45*	-.53

Interpretation: Not interpretable.

Discussion: The highest loadings on this factor are found in the serial relearning of color-forms after the interpolation of a highly interfering color-form list. The high correlation between errors and oscillation creates a kind of artificial "doublet" factor here. Interpolated color-form learning also has substantial correlations with this factor, which suggests that if the factor must be given a label it is probably best described as susceptibility to retroactive and associative interference in serial learning. Presumably there is some RI even in the IL-Low condition (interpolated trigram list), but it does not show up on this factor. This makes its description as retroactive interference somewhat dubious. The situation would have been much improved if two or more tests of RI (with highly interfering interpolated learning) in serial learning had been included in this battery. It is clear that relearning after a highly interfering interpolated list involves a different factor than is involved in original learning, but there is not enough information here to afford a psychological interpretation of this factor.

Factor F

	Variable	Loading	Correlation
(06)	Immediate Digit Span 4-7	-.79*	-.83
(33)	Age	.72*	.65
(05)	Digit Span: Proactive Inhibition	-.68*	-.69
(04)	Delayed Digit Span 4-7	-.65*	-.71
(01)	Immediate Digit Span 2-9	-.22	-.49
(02)	Delayed Digit Span 2-9	-.15	-.38
(09)	Digit Span 5-12	-.16	-.40
(10)	Repeated Digit Span 5-12	-.18	-.43

Interpretation: Weak initial registration of the stimulus trace.

Discussion: The digit span paradigm involving proactive inhibition (PI) has its largest loading on this factor. RI correlates only .23 with this factor. It is not called a proactive inhibition factor because other tests are loaded on it as well. These are the memory span tests which involve relatively little interference of the type represented by Factor A. The digit series of 4 to 7 items in length involve comparatively little intralist retroactive inhibition. IDs in this factor play a part in all memory span performance but it is a small part when conditions are such as to bring Factor A strongly into the picture. The important point is that IDs in PI and in relatively short digit series are determined by a different factor (Factor F) than IDs in RI and supraspan series (Factor A). Factor F is interpreted here as weak initial registration of the stimulus trace, leaving less trace to be consolidated. In the PI paradigm the persisting trace of List 1 presumably weakens the registration of List 2.

It is interesting to note that age has its highest loading on this factor and has only a small loading on Factor A. If this finding holds up in further studies it would suggest that the age decrement in learning and short term memory is due mainly to increasingly poor registration of the stimulus input with increasing age rather than with increased susceptibility to interference with consolidation. Since at least two factors (A and F) are involved in STM, it could well be the case that the lower memory span of childhood and the lower span of old age have quite different causes. Further investigation of the relationship of age to Factors A and F will, of course, require sampling from a much greater age range than exists in the present sample.

Factor G

Variable	Loading	Correlation
+(19) Serial RI, RL-L (errors)	.97*	.88
+(25) Serial RI, RL-L (oscillation)	.95*	.91
+(30) Serial Words (oscillation)	.33	-.43

Interpretation: Artificial "doublet" factor specific to relearning of color-from serial list after a low interference interpolated list (trigrams).

Discussion: Because of the high correlation between errors and oscillation, this one test has become isolated as a single factor, but since no other tests have appreciable loadings on this factor it is regarded as a doublet, or artificial factor. In an orthogonal rotation this doublet variance would be lost to further analysis, since by definition it would be uncorrelated with other factors. With oblique rotation, however, there is a possibility that this factor will be correlated with others and we might therefore be able to discern the factorial nature of this test. But the most we can say at the moment is that relearning of a serial list after low interpolated interference stands somewhat apart from original learning of the list and from relearning after high interpolated interference.

Factor H

Variable	Loading	Correlation
+(23) Serial RI, IL-L (oscillation)	1.00*	.77
+(17) Serial RI, IL-L (errors)	.86*	.79
+(31) Serial, Sum of Trigrams (oscil.)	.42*	.70
+(30) Serial, Words (oscillation)	.41*	.58
(33) Age	-.40	-.41
+(22) Serial RI, OL-H (oscillation)	.19	.44
+(28) Serial, Sum of Trigrams (errors)	.15	.50

Interpretation: Not clear, but may be tentatively described as resistance to oscillation in serial learning of verbal materials.

Discussion: The errors and oscillation index on the single test (interpolated trigram list in the Serial RI paradigm) creates a kind of artificial doublet here, but it carries a bit more meaning than most doublets to the extent that a few independent tests are also substantially loaded on the factor or have sizable correlations with the factor. All but one of these tests involve verbal materials, and oscillation is more heavily loaded than the error score.

Age has a negative correlation with this factor, which might cause one to wonder if this factor bears any resemblance to Factor F, on which age had its highest loading. The correlation between Factors F and H, however, is only -.01 (see Table 29).

Factor I

Variable	Loading	Correlation
+(21) Serial RI, OL-L (oscillation)	-.97*	-.79
+(24) Serial RI, IL-H (oscillation)	-.69*	-.57
+(32) Serial, Sum of CF (oscillation)	-.69*	-.75
+(15) Serial RI, OL-L (errors)	-.67*	-.67
+(28) Serial, Sum of Trigrams (errors)	-.43	-.69
(07) Immediate Color-form Span 2-9	.41	.10
(09) Digit Span 5-12	-.17	-.42
(10) Repeated Digit Span 5-12	-.11	-.42
+(27) Serial Words (errors)	-.28	-.53
+(31) Serial Sum of Trigrams (oscil.)		-.47
+(29) Serial, Sum of CF (errors)		-.38

Interpretation: Susceptibility to intralist interference and oscillation in serial learning.

Discussion: This factor has its largest loadings on the color-form lists (which are presumed to produce greater intralist interference than words or trigrams) and particularly on the index of oscillation, which is presumed to reflect interference tendencies. This seems to be a genuine factor and is not the result of a doublet created by the correlation between the error and oscillation measures on any single test.

The fact that some of the digit span tests correlate with this factor suggests that we may be dealing here with interference with consolidation similar to that hypothesized for Factor A. The great degree of separation of memory span and serial learning in the first order factors could have resulted from many procedural differences in the two forms of learning, such as a 1 sec. presentation rate in the digit series tests vs. a 3 sec. rate for the serial lists, and so forth. Though Factor I is the clearest and most substantial of the first-order factors involving serial learning, it still leaves a good deal to be desired from the standpoint of interpretation.

Factor J

Variable	Loading	Correlation
(44) Progressive Matrices	.81*	.66
+(36) Stroop Color-Speed Factor	.77*	.79
+(30) Serial Words (oscillation)	-.34	-.49

Interpretation: Not interpretable.

Discussion: This odd factor seems to consist of the speed of color-naming (with speed of word reading partialed out) and whatever is measured by the Progressive Matrices which is not absorbed by the other factors. Since the Matrices Test has its highest loading on this factor, it is suggested that in the present group of Ss the main source of variance on the Matrices is probably not the "g" factor of intelligence. The fact that oscillation in serial learning of words correlates .49 with this factor does not help our understanding.

Factor K

	Variable	Loading	Correlation
+(35)	Stroop Interference Factor	-1.00*	-.73
+(34)	Stroop Speed Factor	-.76*	-.72
(43)	Grade Point Average	-.61*	-.46
(02)	Delayed Digit Span 2-9	-.13	-.44
+(27)	Serial Words (errors)	-.13	-.52
+(28)	Serial, Sum of Trigrams (errors)	-.36	-.54
+(30)	Serial Words (oscillation)	-.17	-.54
+(31)	Serial, Sum of Trigrams (oscil.)	-.34	-.47

Interpretation: Susceptibility to response competition interference.

Discussion: The Stroop Interference score, with its loading of -1.00, virtually defines this factor. The Stroop Interference score seems to reflect in almost pure form response competition between two habits of unequal strengths (word reading and color naming). The Stroop Speed score, which is experimentally independent of the Interference score, is also loaded on this factor, suggesting that sheer speed of word reading is determined in part by response competition. The interpretation of this factor as more than just a Stroop factor seems justified by the substantial correlations of other tests with this factor. The fact that IL-H (high interfering interpolated list) of the Serial RI paradigm, on which one would expect considerable response competition, does not show up prominently on Factor K does not seriously weaken the interpretation of this factor as susceptibility to response competition. For we see that a large part of the variance in IL-H has been absorbed by Factors I and E, and we shall see that Factor K coalesces with Factors I and E in the second-order factor analysis.

Interference due to response competition is seen as a different kind of interference than that represented by Factor A, which is interference with the process of trace consolidation. Response competition interference occurs after traces have already been consolidated to some degree, and the response competition may also come from associations acquired outside the experimental situation. It may be for this reason that the Word and Trigram serial lists figure rather prominently on this factor. The pattern of correlations in Table 28 (comparing Factors A and K) suggests that the color-form serial lists are somewhat more influenced by intralist interference with consolidation (Factor A), while the word and trigram lists are more influenced by response competition interference (Factor K). This may account for the lack of high correlations between serial learning of color-forms and serial learning of words and trigrams. Previous experiments have suggested that color-forms are almost totally free of extra-experimental associations, while just the opposite is known to be true for words and high frequency trigrams such as were used in the present study.

Finally, we notice that college Grade Point Average has its highest loading on this response competition factor. This seems to make sense, if we assume that much of the GPA is based on examinations in courses and that one source of difficulty in examinations is interference or competition among already learned or consolidated responses or associations. Also, consistent with the Spence-Taylor hypothesis concerning the relationship between drive or anxiety (measured by the Neuroticism scales in the present study) and level of performance as a function of the degree of opportunity for the occurrence of response competition is the fact that GPA has a loading of -.31 on Factor D (Neuroticism), along with the loading of -.61 on Factor K (response competition).

Factor L

	Variable	Loading	Correlation
(13)	Backward Digit Span, Pre-cued	-.91*	-.80
(14)	Backward Digit Span, Post-cued	-.82*	-.69
+(29)	Serial, Sum of CF (errors)	-.43	-.41
(01)	Immediate Digit Span 2-9	-.24	-.51
(05)	Digit Span: Proactive Inhibition	-.09	-.40
(06)	Immediate Digit Span 4-7	-.21	-.40
(09)	Digit Span 5-12	-.16	-.51
(10)	Repeated Digit Span 5-12	-.21	-.56
(12)	Forward Digit Span, Post-cued	-.19	-.49
+(27)	Serial Words (errors)	-.08	-.50
+(30)	Serial Words (oscillation)	-.33	-.56
+(32)	Serial, Sum of CF (oscillation)	-.36	-.57

Interpretation: Poor backward memory span.

Discussion: This factor is very clear, although its psychological nature is not understood. When Ss are asked how they recall a digit series in the backward direction, most Ss report that what they actually do is something quite different from recalling the list in the forward direction. They seem to rehearse the series, or some part of the series, in the forward direction to keep it in mind while "jumping" to the end of the series to "read out" the last item, then the next to the last, and so on, all the while trying to keep the earlier items in mind. These items are reported to be "held in mind" in a forward direction. The factor analysis indicates that all this involves a different process than occurs in the other tests in this battery. Many of the other tests, however, have substantial correlations with this factor, as can be seen in the listing above. Serial learning as well as a number of the forward memory span tasks are correlated with the factor. Factor L (poor backward memory span) is not entirely independent of other factors. As can be seen in Table 29, it is correlated -.41 with Factor A (resistance to interference) and .51 with Factor K (susceptibility to response competition).

What has become of the Pre-cue vs. Post-cue variable in our factor analysis? This experimental variable apparently does not involve any separate factor. Though Pre and Post cuing make for a large difference in performance level, this experimental variable involves no additional source of ID variance. Pre and Post-cued memory span have essentially the same factorial composition, but the pre-cued series have somewhat higher loadings on the salient factors than the post-cued series.

3. Rotated Oblique Second-Order Factors.

We now factor analyze the first-order factors to obtain the second-order oblique factors. The intercorrelations of the first-order factors are shown in Table 29. The rotated oblique second-order factors are presented in Table 30 and the correlations of the variables with these factors are given in Table 31. Four significant factors were extracted from the correlation matrix shown in Table 29; these factors account for 40 percent of the variance in this matrix. As before, interpretations of the factors is based on tests with loadings greater than .40. Since the second-order factors are much more general than the more specific first-order factors, we should expect many more of the tests to show appreciable loadings on the second-order factors than on those of the first order.

Factor A'

Variable		Loading	Correlation
+(30)	Serial Words (oscillation)	.70*	.67
(38)	MPI-Extraversion	.61*	.61
+(27)	Serial Words (errors)	.56*	.61
(42)	EPI, Form B, Extraversion	.56*	.54
(45)	Time for Progressive Matrices	.51*	.51
+(34)	Stroop, Speed Factor	.49*	.55
+(28)	Serial, Sum of Trigrams (errors)	.45*	.66
+(35)	Stroop, Interference Factor	.43*	.44
(41)	EPI, Form A, Extraversion	.43	.48

Interpretation: Not clear, but might be called resistance to response competition.

Discussion: This factor is such an odd mixture that it is difficult to interpret. Tests involving response competition and possible extra-experimental interference, such as the word and trigram serial learning, are loaded on this factor. Extraversion seems to manifest itself in the learning domain largely through this factor. An adequate psychological characterization of the factor seems hardly possible with the present information. Most of the variance in Factor A' is drawn from the correlation of -.30 between the first-order Factors B (Extraversion) and K (susceptibility to response competition). Part of the genotype of extraversion might involve resistance to response competition, but the writer knows of no theoretical basis for this conjecture.

There can be little doubt that Extraversion is in some way involved in serial learning. In another study, which is too involved to describe in detail here, serial learning protocols of 60 Ss were scored in a variety of ways yielding 19 experimentally interdependent scores (e.g. trials to criterion, errors, omissions, oscillation, length of forward and backward remote associations, index of skewness, index of bowing, etc.). A principal components analysis

Table 29

Intercorrelations Among the Rotated Oblique Factors[1] (Table 27)

(Decimals omitted)

	A	B	C	D	E	F	G	H	I	J	K	L
A												
B	-06											
C	-26	10										
D	23	-06	-10									
E	-05	09	26	-07								
F	-26	-16	00	-04	-15							
G	06	12	-20	-11	-06	-04						
H	29	05	-35	09	00	-01	02					
I	-26	-19	26	-26	-09	19	-08	-43				
J	-22	-10	20	-13	08	-05	05	-20	12			
K	-33	-30	-05	-17	-23	13	15	-26	35	12		
L	-41	-13	-05	-24	-07	29	14	-22	38	03	51	

[1] Identifying labels of the factors are given below:

A. Resistance to retroactive interference in short term memory, or resistance to interference with consolidation of the memory trace.

B. Extraversion.

C. Not interpreted.

D. Neuroticism.

E. Not interpreted.

F. Weak initial registration of the stimulus trace.

G. Not interpreted (artificial doublet).

H. Resistance to oscillation (or interference) in serial learning of verbal materials.

I. Susceptibility to intralist interference and oscillation in serial learning.

J. Not interpreted.

K. Susceptibility to response competition interference.

L. Poor backward memory span.

Table 30

Rotated Oblique Second-Order Factor Coefficients (Decimals omitted)

Variable		A'	B'	C'	D'
1. IDS, 2-9	(01)	23	22	16	58
2. DDS, 2-9	(02)	29	04	19	45
3. DS-RI, 4-7	(03)	-03	-06	-12	58
4. DDS, 4-7	(04)	-21	17	12	79
5. DS-PI, 4-7	(05)	-03	18	05	73
6. IDS, 4-7	(06)	-13	27	16	74
7. ICFS, 2-9	(07)	-07	-40	-02	25
8. DCFS, 2-9	(08)	-08	-28	-02	34
9. DS, 5-12	(09)	-03	-07	04	79
10. RDS, 5-12	(10)	-02	-09	03	82
11. FDS-Pre	(11)	-14	-13	-10	66
12. FDS-Post	(12)	14	14	-28	65
13. BDS-Pre	(13)	12	02	04	55
14. BDS-Post	(14)	13	-19	05	47
15. Ser. RI, OL-L (Errors)	(15)*	03	-44	44	07
16. " " OL-H "	(16)*	-11	-40	22	17
17. " " IL-L "	(17)*	18	-28	46	05
18. " " IL-H "	(18)*	-37	-69	18	22
19. " " RL-L "	(19)*	-04	-17	66	01
20. " " RL-H "	(20)*	-15	-38	07	-09
21. " " OL-L (Oscil.)	(21)*	24	-16	15	12
22. " " OL-H "	(22)*	24	-62	04	-20
23. " " IL-L "	(23)*	19	-11	26	03
24. " " IL-H "	(24)*	-11	-55	34	02
25. " " RL-L "	(25)*	-10	03	76	09
26. " " RL-H "	(26)*	07	05	-33	09
27. Ser. Words (Errors)	(27)*	56	05	-25	45
28. " Sum of Tri "	(28)*	45	-23	44	16
29. " Sum of CF "	(29)*	03	-25	35	32
30. " Words (Oscil)	(30)*	70	-33	-42	09
31. " Sum of Tri "	(31)*	38	-56	07	01
32. " Sum of CF "	(32)*	22	-15	02	46
33. Age (months)	(33)	-23	-15	-20	-30
34. Stroop, Speed Factor	(34)*	49	-10	-08	21
35. " Interference Factor	(35)*	43	06	-02	09
36. " Color-Speed Factor	(36)*	-15	55	03	27
37. MPI-Neuroticism	(37)	04	-17	-23	31
38. MPI-Extraversion	(38)	61	24	36	-17
39. EPI, Form A, Neuroticism	(39)	-01	-16	-31	38
40. EPI, Form B, "	(40)	-07	-30	-27	26
41. EPI, Form A, Extraversion	(41)	43	-04	47	-27
42. EPI, Form B "	(42)	56	10	25	-23
43. College Grade Point Average	(43)	07	37	-03	41
44. Progressive Matrices	(44)	-46	-02	13	25
45. Total time for P.M.	(45)*	51	49	09	23

* Asterisks indicate that the variable has been reversed in sign, so that all the memory span, serial learning, and Stroop tasks are keyed in the same direction representing "goodness" of performance.

Table 31

Correlations of Variables With Oblique Second-Order Factors

(Decimals Omitted)

Variable		A	B	C	D
1. IDS,2-9	(01)	40	-07	30	57
2. DDS,2-9	(02)	47	-21	32	54
3. DS-RI,4-7	(03)	12	-32	-07	58
4. DDS,4-7	(04)	03	-14	16	67
5. DS-PI,4-7	(05)	16	-13	13	64
6. IDS,4-7	(06)	08	-03	22	60
7. ICFS,2-9	(07)	08	-51	-03	41
8. DCFS,2-9	(08)	07	-41	-02	44
9. DS,5-12	(09)	22	-41	12	82
10. RDS,5-12	(10)	24	-44	11	86
11. FDS-Pre	(11)	05	-40	-08	66
12. FDS-Post	(12)	23	-19	-16	60
13. BDS-Pre	(13)	29	-25	13	58
14. BDS-Post	(14)	31	-42	12	60
15. Ser. RI, OL-L (Errors)	(15)*	26	-44	43	32
16. " " OL-H "	(16)*	07	-44	18	34
17. " " IL-L "	(17)*	37	-31	49	27
18. " " IL-H "	(18)*	-13	-70	06	44
19. " " RL-L "	(19)*	17	-12	64	15
20. " " RL-H "	(20)*	-09	-31	00	04
21. " " OL-L (Oscil.)	(21)*	35	-24	22	28
22. " " OL-H "	(22)*	31	-57	04	15
23. " " IL-L "	(23)*	29	-14	31	17
24. " " IL-H "	(24)*	09	-51	28	27
25. " " RL-L "	(25)*	13	06	75	13
26. " " RL-H "	(26)*	00	-02	-30	05
27. Ser. Words (Errors)	(27)*	61	-27	-04	57
28. Ser. Sum of Tri "	(28)*	66	-36	57	44
29. " " of CF "	(29)*	26	-38	38	48
30. " Words (Oscil.)	(30)*	67	-53	-24	39
31. " Sum of Tri "	(31)*	51	-63	14	37
32. " Sum of CF "	(32)*	38	-39	12	59
33. Age (months)	(33)	-35	02	-31	-33
34. Stroop, Speed Factor	(34)*	55	-29	07	39
35. " Interference Factor	(35)*	44	-06	11	19
36. " Color-Speed Factor	(36)*	-17	46	05	-02
37. MPI-Neuroticism	(37)	10	-33	-20	37
38. MPI-Extraversion	(38)	61	22	53	-05
39. EPI, Form A, Neuroticism	(39)	05	-35	-28	42
40. EPI, Form B, "	(40)	-02	-42	-28	38
41. EPI, Form A, Extraversion	(41)	48	03	56	-08
42. EPI, Form B, "	(42)	54	12	39	-09
43. College Grade Point Average	(43)	11	18	06	26
44. Progressive Matrices	(44)	-35	-04	03	14
45. Total time for P.M.	(45)*	51	30	28	17

* Asterisks indicate that the variable has been reversed in sign, so that all the memory span, serial learning, and Stroop tasks are keyed in the same direction representing "goodness" of performance.

reduced these 19 derived scores to five independent dimensions. The MPI-Extraversion scale was included in the analysis and was found to have a communality of .48, which means that if all the serial learning derived scores were put into a multiple regression equation, their multiple R with Extraversion would not be less than .69. In other words, nearly half the variance (48 percent to be exact) in Extraversion would have been predictable from performance in a single serial learning task (a nine-item color-form list presented at a 3 sec. rate). In still another study (Jensen, 1962) Extraversion was found to be positively correlated with speed of serial learning when the list (color-forms) was presented at a 2 sec. rate. The correlation between serial learning and extraversion is not understood in theoretical terms, but these findings make it important that measures of extraversion be included in future studies of IDs in serial learning.

Factor B'

Variable	Loading	Correlation
+(18) Serial RI, IL-H (errors)	-.69*	-.70
+(22) Serial RI, OL-H (oscillation)	-.62*	-.57
+(31) Serial, Sum of Trigrams (oscil.)	-.56*	-.63
+(24) Serial RI, IL-H (oscillation)	-.55*	-.51
+(36) Stroop Color-speed Factor	.55*	.46
(45) Time for Progressive Matrices	.49*	.30
+(15) Serial RI, OL-L (errors)	-.44	-.44
(07) Immediate Color-form Span 2-9	-.40	-.51
+(16) Serial RI, OL-H (errors)	-.40	-.44
(08) Delayed Color-form Span 2-9	-.28	-.41
(09) Digit Span 5-12	-.07	-.41
(10) Repeated Digit Span 5-12	-.09	-.44
(11) Forward Digit Span, pre-cued	-.13	-.40
(14) Backward Digit Span-Post-cued	-.19	-.42
+(30) Serial Words (oscillation)	-.33	-.53
(40) EPI, Form B, Neuroticism	-.30	-.42

Interpretation: Susceptibility to interference with consolidation in serial learning.

Discussion: This factor of interference in serial learning stands apart from what is genotypically probably the same type of interference found in the memory span tests because of the considerable procedural differences between the serial learning and memory span tasks, involving such parameters as rate of presentation, form of recall, etc. The memory span aspect of this factor is Factor D', and it is intersting to note that Factors B' and D' correlate .45, which is the highest correlation among any of the second order factors. Factor B' involves intratask retroactive inhibition of memory traces in the course of serial learning. Unlike Factor A' it is not thought to involve competition or mutual blocking between more or less consolidated response tendencies.

It should be pointed out that no very clear distinction can be made in the present study between interference with consolidation and speed of consolidation. The more rapidly a trace is consolidated, the more rapidly it becomes resistant to interference. The interference factors involving consolidation could also be called speed of consolidation. There is some reason for believing, however, that speed of consolidation and resistance to interference are not one and the same thing even though they are not actually separated in the present factor analyses. Delayed Digit Span 4-7 and Digit Span-RI 4-7, for example, correlate .80 after correction for attenuation (.48 without correction). These two tests are formally very similar except for the degree of interference interpolated between the presentation and recall of the digit series. In Delayed Digit Span plus (+) and minus (-) signs are interpolated between presentation and recall; in Digit Series-RI another series of digits (taking up the same amount of time as the pluses and minuses) is interpolated between presentation and recall. If speed of consolidation were the only factor involved here, it seems one should expect a higher correlation and as a corollary one would also expect the two measures to stick together in the factor analyses, which they do not. It is suggested that speed of consolidation and resistance to interference with consolidation are not the same thing from the standpoint of IDs, but they have not been clearly separated in the present study. An experiment designed to disentangle these two aspects of consolidation (i.e., speed of consolidation and resistance to interference with consolidatiom) is now being carried on in the writer's laboratory.

Factor C'

	Variable	Loading	Correlation
+(25)	Serial RI, RL-L (oscillation)	.76*	.75
+(19)	Serial RI, RL-L (errors)	.66*	.64
(41)	EPI, Form A, Extraversion	.47*	.56
+(17)	Serial RI, IL-L (errors)	.46*	.49
+(15)	Serial RI, OL-L (errors)	.44*	.43
+(28)	Serial, Sum of Trigrams (errors)	.44	.57
+(30)	Serial Words (oscillation)	.42	.24
(38)	MPI-Extraversion	.36	.53

Interpretation: Not interpretable.

Discussion: This appears to be mainly determined by a doublet created by the high correlation between errors and oscillation on the Serial RL-L test.

Relearning, particularly relearning after a high level of interpolated interference (i.e., Serial RI, RL-H condition), is hardly accounted for by the variance in the present battery. RL after high interpolated interference (i.e., RL-H) appears only as a doublet (Factor E) in the first order factor analysis and practically vanishes at higher levels of analysis. The highest correlation that RL-H has with anything else in the matrix of zero order correlations (Table 25), aside from the correlation between errors and oscillation index on this test, is .38 (IL-H, oscillation). The lack of correlation between OL and RL will have to be studied in its own right. Other studies have shown a lack of correlation between speed of OL and measures of retention when there has been no experimentally interpolated interference (Underwood, 1954; Stroud & Schoer, 1959). But it comes as something of a surprise to find a negative correlation (though not significantly greater than zero) between OL and RL after high interpolated interference. It seems fairly safe to conclude that IDs in relearning after original learning has been severely retroactively inhibited involves a different genotype than is involved in IDs in OL. Before this conclusion can be made any more firm, however, we will have to know more about the reliability of IDs in relearning and about the effects of practice on the factorial composition of serial learning. The problem of the factorial nature of OL, IL, and RL in serial learning calls for a large-scale experiment in its own right. It would seem rather foolhardy to go on to investigating IDs in other forms of learning before we have even begun to understand IDs in serial learning, which seems relatively simple compared with many other forms of human learning.

A lack of correlation, such as we find between OL and RL, is a challenging puzzle to experimental and differential psychology. Significant positive correlations among IDs in what appear to be formally similar tasks are the usual and expected thing. It is when such correlations fail to materialize that we should sit up and take notice. Advancement of our understanding of IDs in learning will depend in large part upon thorough investigation of these salient, surprise phenomena wherever they may turn up in the course of research.

Factor D'

	Variable	Loading	Correlation
(10)	Repeated Digit Span 5-12	.82*	.86
(04)	Delayed Digit Span 4-7	.79*	.67
(09)	Digit Span 5-12	.79*	.82
(06)	Immediate Digit Span 4-7	.74*	.60
(05)	Digit Span: Proactive Inhibition	.73*	.64
(11)	Forward Digit Span, Pre-cued	.66*	.66
(12)	Forward Digit Span, Post-cued	.65*	.60
(01)	Immediate Digit Span 2-9	.58*	.57
(03)	Digit Span: Retroactive Inhibition	.58*	.58
(13)	Backward Digit Span, Pre-cued	.55*	.58
(14)	Backward Digit Span, Post-cued	.47*	.60
+(32)	Serial, Sum of CF (oscillation)	.46*	.59
(02)	Delayed Digit Span 2-9	.45*	.54
(27)	Serial Words (errors)	.45	.57
(43)	Grade Point Average	.41*	.26
(07)	Immediate Color-form Span 2-9	.25	.41
(08)	Delayed Color-form Span 2-9	.34	.44
+(18)	Serial RI, IL-H (errors)	.22	.44
+(28)	Serial, Sum of Trigrams (errors)	.16	.44
+(29)	Serial, Sum of CF (errors)	.32	.48
(39)	EPI, Form A, Neuroticism	.38	.42

Interpretation: Speed of consolidation and resistance to interference with consolidation.

Discussion: Repeated Digit Span, which has the highest loading and highest correlation with this factor, was hypothesized to tap consolidation, and where marked interference is involved in the task the factor loadings are somewhat lower. This rather broad factor is therefore believed to represent mainly the speed of consolidation and to a lesser degree resistance to interference with consolidation. It could be called a general memory span factor, but this label tends to conceal the idea that IDs in memory span are a result of three correlated first-order factors: strength of registration of the stimulus trace, speed of consolidation, and resistance to interference with consolidation. Factor D' is the common element in these first order factors. A number of the serial learning tasks have moderate correlations with this factor, though most of the serial tasks are more heavily loaded on Factors A' and B', both of which, however, have substantial correlations with Factor D', as shown in Table 32.

4. Rotated Oblique Third-Order Factors.

Table 32 shows the correlations among the second-order factors which provided the basis for the third-order factor analysis. At this level two very broad factors emerge. Together they account for 28.5 percent of the variance in the matrix of Table 32. Thus, of the 77 percent of the total variance which is accounted for by the whole factor analysis, 37 percent is taken up by the first-order factors, 11.5 percent by the second-order factors, 16.5 percent by the third-order factors, and 12 percent by the fourth-order factor, which is the general factor in this battery of tests.

Table 33 shows the third-order factor coefficients (loadings) and the correlations of the variables with the factors.

Table 32

Correlations Among the Second-Order Factors[1] (Table 31) (Decimals omitted)

	A'	B'	C'	D'
A'				
B'	-18			
C'	27	06		
D'	29	-45	11	

---

[1] Identifying labels of the factors are given below:

A'. Not clear, possibly resistance to response competition.

B'. Susceptibility to interference with consolidation in serial learning.

C'. Not interpreted.

D'. Speed of consolidation and resistance to interference with consolidation.

Table 33

Rotated Oblique Third-Order Factor Coefficients and Correlations of Variables With Third Order Factors (Decimals Omitted)

Variable		Coefficients		Correlations	
		A"	B"	A"	B"
1. IDS,2-9	(01)	28	40	42	50
2. DDS,2-9	(02)	36	41	50	53
3. DS-RI,4-7	(03)	54	-11	51	08
4. DDS,4-7	(04)	39	08	42	21
5. DS-PI,4-7	(05)	40	12	44	26
6. IDS,4-7	(06)	27	19	33	29
7. ICFS,2-9	(07)	56	-15	51	04
8. DCFS,2-9	(08)	51	-12	47	05
9. DS,5-12	(09)	68	05	70	28
10. RDS,5-12	(10)	72	04	74	29
11. FDS,Pre	(11)	64	-18	58	04
12. FDS-Post	(12)	51	-09	48	09
13. BDS-Pre	(13)	45	16	50	31
14. BDS-Post	(14)	57	11	61	30
15. Ser. RI, OL-L (Errors)	(15)*	34	32	45	43
16. " " OL-H "	(16)*	41	03	42	17
17. " " IL-L "	(17)*	22	46	38	54
18. " " IL-H "	(18)*	65	-25	57	-02
19. " " RL-L "	(19)*	-03	54	15	53
20. " " RL-H "	(20)*	21	-14	17	-07
21. " " OL-L (Oscil.)	(21)*	26	26	35	35
22. " " OL-H "	(22)*	47	00	47	16
23. " " IL-L "	(23)*	11	34	22	37
24. " " IL-H "	(24)*	39	09	42	23
25. " " RL-L "	(25)*	-20	66	03	60
26. " " RL-H "	(26)*	14	-24	06	-19
27. Ser. Words (Errors)	(27)*	54	17	60	36
28. " Sum of Tri "	(28)*	35	65	57	77
29. " Sum of CF "	(29)*	40	30	50	43
30. " Words (Oscil.)	(30)*	69	-03	68	20
31. " Sum of Tri "	(31)*	60	16	66	36
32. " Sum of CF "	(32)*	56	14	61	33
33. Age (months)	(33)	-11	-40	-25	-44
34. Stroop, Speed Factor	(34)*	42	22	50	37
35. " Interference Factor	(35)*	15	28	24	33
36. " Color-Speed Factor	(36)*	-34	10	-30	-02
37. MPI-Neuroticism	(37)	48	-21	40	-05
38. MPI-Extraversion	(38)	-28	78	-02	69
39. EPI, Form A, Neuroticism	(39)	54	-32	43	-13
40. EPI, Form B, "	(40)	54	-37	41	-18
41. EPI, Form A, Extraversion	(41)	-20	68	04	62
42. EPI, Form B, "	(42)	-20	61	01	54
43. College Grade Point Average	(43)	01	15	06	16
44. Progressive Matrices	(44)	06	-16	00	-14
45. Total time for P.M.	(45)*	-15	56	04	51

* Asterisks indicate that the variable has been reversed in sign, so that all the memory span, serial learning, and Stroop tasks are keyed in the same direction representing "goodness" of performance.

Factor A"

Variable	Loading	Correlation
(10) Repeated Digit Span 5-12	.72*	.74
+(30) Serial Words (oscillation)	.69*	.68
(09) Digit Span 5-12	.68*	.70
(18) Serial RI, IL-H (errors)	.65*	.57
+(31) Serial, Sum of Trigrams (oscil.)	.60*	.66
(14) Backward Digit Span, Post-cued.	.57*	.61
(07) Immediate Color-form Span 2-9	.56*	.51
+(32) Serial, Sum of CF (oscillation)	.56*	.61
(03) Digit Span: Retroactive Inhibition	.54*	.51
+(27) Serial Words (errors)	.54*	.60
(39) EPI, Form A, Neuroticism	.54*	.43
(40) EPI, Form B, Neuroticism	.54*	.41
(08) Delayed Color-form Span 2-9	.51*	.47
(12) Forward Digit Span, Post-cued	.51*	.48
(37) MPI-Neuroticism	.48*	.40
+(22) Serial RI, CL-H (oscillation)	.47*	.47
(13) Backward Digit Span, Pre-cued	.45*	.50
+(34) Stroop Speed Factor	.42*	.50
+(16) Serial RI, CL-H (errors)	.41*	.42
(05) Digit Span: Proactive Inhibition	.40*	.44
+(29) Serial, Sum of CF (errors)	.40*	.50

Interpretation: Learning ability, free of extraversion.

Discussion: This seems to be the first factor that represents what might be called a general factor of learning ability. But note that none of the tests involving relearning appears on this factor. It is somewhat less general than the next, fourth-order factor, because the rotation to approximate simple structure has excluded extraversion from this factor. Factor A" is correlated +.34 with Factor B", which is identified as Extraversion as it is manifested in the learning domain.

Factor B"

Variable	Loading	Correlation
(38) MPI-Extraversion	.78*	.69
(41) EPI, Form A, Extraversion	.68*	.62
+(25) Serial RI, RL-L (oscillation)	.66*	.60
+(28) Serial, Sum of Trigrams (errors)	.65*	.77
(42) EPI, Form B, Extraversion	.61*	.54
+(45) Time for Progressive Matrices	.56*	.51
+(19) Serial RI, RL-L (errors)	.54*	.53
+(17) Serial RI, IL-L (errors)	.46*	.54
(02) Delayed Digit Span 2-9	.41*	.53
(01) Immediate Digit Span 2-9	.40*	.50
(33) Age	-.40*	-.44

Interpretation: Extraversion as manifested in rote learning and short-term memory.

Discussion: There is little that can be said about this factor other than what meets the eye in the loadings and correlations listed above. The genotypic mechanism that links extraversion with these various measures of learning and STM is not understood. The fact that extraversion correlates positively with general learning ability of the kind represented by these tasks is rather puzzling in view of Eysenck's theory of extraversion as involving greater susceptibility to internal inhibition in the Pavlovian sense (see Eysenck, 1965). Could it be that extraverts are less susceptible to external inhibition, which might be an analogue of the response competition interference hypothesized to play a part in serial learning, particularly in cases where there is a good deal of associative interference due to associations acquired outside the laboratory learning situation?

5. General Factor.

Since Factors A'' and B'' are correlated .34, they yield a single fourth order factor, which is the general factor of this battery. It correlates .72 with the first principal component (Table 26). The loadings of the fourth order general factor are presented in Table 34. This factor accounts for 12 percent of the variance. It is probably best interpreted as general learning ability. It is interesting to note that all of the memory span and serial learning tasks have fairly substantial loadings on this general factor, with the one exception of RL-H, i.e., relearning a color-form list after it has been retroactively inhibited by the learning of an interpolated color-form list. RL-H has a loading on the general factor of .06 for errors and .08 for oscillation. The point is reiterated that IDs in relearning after OL has been retroactively inhibited involve some source of variance that is not represented to any appreciable extent in this whole battery of tests. The internal consistency reliability of total errors in RL-H has been estimated as .90 and the test-retest reliability has been estimated as .32. Using the latter estimate to correct the loading of RL-H on the general factor for attenuation gives a loading of only .11. Are we to conclude that relearning does not involve the same ability as original learning? This may be the case. Another possibility might be that fast learners in OL have consolidated relatively little before the interpolated list has to be contended with, and they therefore suffer greater retroactive inhibition than do slower learners who may actually overlearn much of the serial list during OL. Further investigation of this problem will require methods for equating response strength of OL in fast and slow learners just prior to the presentation of the interpolated list. Merely having all Ss learn the OL list to the same criterion, as was done in the present study, is not an adquate control for a more refined investigation of this problem.

VI. Summary, Conclusions, and Implications

Since detailed discussion of the results has been presented at every step of the way, the present section will focus only on some of the most important findings and their psychological implications.

We began with the question of whether IDs in susceptibility to interference effects in a variety of phenotypically different learning paradigms have a

Table 34

Fourth-Order (General) Factor (Decimals Omitted)

Variable		g
1. IDS,2-9	(01)	56
2. DDS,2-9	(02)	63
3. DS-RI,4-7	(03)	36
4. DDS,4-7	(04)	38
5. DS-PI,4-7	(05)	43
6. IDS,4-7	(06)	38
7. ICFS,2-9	(07)	33
8. DCFS,2-9	(08)	32
9. DS,5-12	(09)	60
10. RDS,5-12	(10)	63
11. FDS-Pre	(11)	38
12. FDS-Post	(12)	35
13. BDS-Pre	(13)	50
14. BDS-Post	(14)	56
15. Ser. RI, OL-L (Errors)	(15)*	54
16. " " OL-H "	(16)*	36
17. " " IL-L "	(17)*	56
18. " " IL-H "	(18)*	33
19. " " RL-L "	(19)*	42
20. " " RL-H "	(20)*	06
21. " " OL-L (Oscil.)	(21)*	43
22. " " OL-H "	(22)*	38
23. " " IL-L "	(23)*	36
24. " " IL-H "	(24)*	40
25. " " RL-L "	(25)*	38
26. " " RL-H "	(26)*	08
27. Ser. Words (Errors)	(27)*	58
28. " Sum of Tri "	(28)*	81
29. " Sum of CF "	(29)*	57
30. " Words (Oscil.)	(30)*	54
31. " Sum of Tri "	(31)*	62
32. " Sum of CF "	(32)*	58
33. Age (months)	(33)	-42
34. Stroop, Speed Factor	(34)*	53
35. " Interference Factor	(35)*	35
36. " Color-Speed Factor	(36)*	-20
37. MPI-Neuroticism	(37)	21
38. MPI-Extraversion	(38)	41
39. EPI, Form A, Neuroticism	(39)	18
40. EPI, Form B, "	(40)	14
41. EPI, Form A, Extraversion	(41)	40
42. EPI, Form B, "	(42)	34
43. College Grade Point Average	(43)	13
44. Progressive Matrices	(44)	-08
45. Total time for P.M.	(45)*	33

* Asterisks indicate that the variable has been reversed in sign.

single common genotype or a number of different genotypes. In an attempt to throw some light on this matter the present study proceeded in two phases. First, a number of short-term memory (STM) and serial learning tasks were tried out on separate groups of Ss under conditions that varied in terms of the amount of interference that had to be overcome in the course of learning and retention. The main sources of experimentally manipulated interference were proactive and retroactive inhibition of STM tasks, retroactive inhibition of serial learning, and interference due to intratask similarity in both STM and serial learning tasks. Those procedures which showed statistically reliable main effects of the experimental variables in the form of performance decrements due to interference, and which also showed a reliable Ss X Experimental Variable interaction, i.e., reliable IDs in the interference effect, were selected for future study by means of factor analysis. A large battery of these STM and serial learning tasks was administered to a single group of Ss for factor analysis. Included in the analysis were a number of other relevant measures: age, grade point average, intelligence test scores (Progressive Matrices), personality measures (Extraversion and Neuroticism), and the Stroop color-word test, which has high fact validity as a measure of IDs in susceptibility to interference due to response competition.

Twelve principal components were extracted from the matrix of intercorrelations; these were rotated to approximate oblique simple structure. The first-order factors were factor analyzed to yield oblique second-order factors; and the same process was repeated until a fourth-order single, general factor was arrived at. The picture of IDs presented here is thus hierarchical. There were 12 first-order factors, accounting for 77 percent of the total variance, though several of these were unclear or uninterpretable. These factors coalesce into 4 second-order factors; these coalesce into 2 third-order factors; and the correlation between these factors results finally in the general factor for this battery.

The specific factors at each level have already been interpreted in the previous section. The main conclusions from these analyses are as follows:

1. Memory span and serial learning have common sources of IDs, as indicated by their similar loadings on the first principal component and by their similar loadings on the fourth-order general factor. Both memory span and serial learning are heavily loaded on what might be called a general factor of learning ability. This is especially interesting in view of the fact that memory span also correlates highly with tests of general intelligence. Most intelligence tests are achievement tests; that is to say, they test the S's store of information and acquired skills. Memory span seems to be the one part of standard intelligence tests, such as the Stanford-Binet and the Wechsler tests, that measures the S's current general learning ability. It is probably little affected by transfer from previous learning and would seem to provide a rather "pure" measure of innate and current learning ability. The memory span tests found in traditional intelligence test batteries could be greatly improved in reliability and in the scope of the abilities that they are capable of tapping. The procedures of the present study indicate some of the ways that memory span tests can be elaborated to yield a more comprehensive sample of the S's learning ability, which seems to be at least a third-order factor. As shown by the first and second levels of our analysis, memory

span is not a unitary ability. Specific diagnosis of learning disabilities might be achieved by means of memory span tasks which are specially designed to tap the main sources of variance in memory span performance.

2. Performance on memory span tasks is not a unitary ability, as shown by the fact that memory span as measured by a variety of means in the present study breaks up into at least three main sources of ID variance. The most important of these factors seems to be resistance to interference with the process of stimulus trace consolidation. There are probably at least two levels of trace consolidation, short-term and long-term. We are dealing here only with short-term consolidation. It is clear that a stimulus trace persists after the external physical stimulus energy has ceased. The persistence or consolidation of this trace can be interfered with by the presentation of other attention-demanding stimuli. The faster the trace is consolidated, the sooner it becomes resistant to external interference. In the present study there is not a clear separation between these two hypothetical processes (speed of trace consolidation and resistance to interference), although it would be difficult to describe the present results in terms of only one or the other of these. It will not be an easy matter to separate these processes experimentally. We have noted however that Factor A in the first-order factor analysis is best identified as a factor of resistance to interference with consolidation, on the grounds that memory span tasks had loadings on this factor which were roughly proportional to the amount of interference that was hypothesized as being involved in the various tasks. Thus, for example, supra-span digit series (i.e., series that exceed the S's memory span as it is traditionally measured) have higher loadings on this interference factor than digit series in the region of the S's memory span. And series that involve high intralist similarity (the color-form series) are more heavily loaded with this factor than series made up of more readily discriminated items, such as digits. When the recall of a digit or color-form series is delayed and an interfering activity is interposed during the delay interval, the scores are more heavily loaded on the interference factor than are scores for immediate recall of the series. And if the interposed material is very similar to the material to be recalled, the factor loading is higher than if the interposed material is dissimilar. The retroactive inhibition paradigm consisted of interpolating another digit series between the presentation of the first, to-be-recalled series and the time for its recall. This test was heavily loaded on the interference factor. Since this factor involves resistance to retroactive inhibition of the trace of any stimulus input by subsequent input, it must obviously play a large part in determining the length of the S's memory span itself, and indeed, all the memory span tasks had substantial correlations with this factor.

One of the most interesting findings of the study was the fact that proactive inhibition of memory span does not involve the same factor which is involved in retroactive inhibition. Though proactive and retroactive inhibition result in about the same average amount of performance decrement in memory span tasks, IDs in these two forms of interference apparently have different genotypes. The proactive inhibition (PI) paradigm had its largest loading on a factor which was hypothesized to involve the strength of initial registration of the stimulus trace. Also loaded on this factor we find memory span tasks in which the length of the series is in the region of the S's memory span. It appears that when a consolidation process is set in motion by the S's attending to a prior digit series, the initial trace of the second, to-be-recalled series is weakened. We cannot say that the trace of the second series is interfered with by the traces of the first series, for this would suggest that proactive inhibition should be highly correlated with retroactive inhibition, which was not found to be the case. Proactive inhibition does not seem to involve interference with the consolidation of the to-be-recalled series. It might be thought that retroactive and proactive inhibition result from these paradigms because the S's attention must be spread over two digit series (the one to be recalled and the interfering series), which would tend to attenuate his retention of either series. But this interpretation would not predict the present finding. If proactaive and retroactive inhibition were just a matter of spreading attention too widely and too thinly, there should be no reason for RI and PI to come out on different factors. It seems preferable to hypothesize different processes and different sources of IDs underlying these two types of interference. Thus, as suggested earlier, STM might be likened to a photographic process. First, there is initial sensitivity of the photographic emulsion, and if the plate is pre-exposed it will be rendered less sensitive. After the image has been registered, it must be developed (consolidated), and the stronger the developing solution, the faster the film becomes immune to interference from unwanted light. The three elements here that determine the final produce (clarity and fidelity of the image, or probability of correct read-out or recall) are thus (a) strength of initial registration, (b) speed of consolidation, and (c) resistance to interference during the course of consolidation. As we have already indicated, b and c are not clearly separated in the first-order factor analysis, for much the same sort of reason that it would be difficult to separate these two factors in the photographic analogy. However, if mere lapse of time between exposure and recall were all that were involved in the probability of correct recall, we should not expect different factor loadings as a function of the kind of material interpolated in the delay interval. Yet we have seen that the degree of similarity between the series to be recalled and the stimuli interpolated between presentation and recall affects the loading of the test on the interference factor (Factor A). Furthermore, the Repeated Memory Span test, which involves a good deal of interference, since most of the series were or supraspan length (5 to 12 items), does not have the highest loading on the interference factor but has the highest loading on the first principal component, which may be regarded as a general learning ability factor, and it also has the highest loadings on the factors at the second and third levels of analysis which are describable as speed of consolidation or short-term learning ability. If the only effect of repeating a digit series was to overcome interference, one would expect this test to behave much like the other memory span tasks throughout the factor analyses; instead it has the highest loadings on what seem to be learning ability factors and comparatively moderate loadings on the interference factor. We are left with the impression that resistance to interference is not the same factor as speed of consolidation, but a clear-cut demonstration of the separateness of these two hypothesized factors awaits further research.

Despite the great restriction of range on the variable of age in the present sample, age had its highest loading on the factor described as strength of initial registration. Older S's, according to this interpretation, have a weaker initial trace. Age did not load appreciably on the interference factor, but it had a substantial negative correlation with the general factor. This finding is of interest to the psychology of aging and should be looked into further. One of the tasks of the psychology of aging is to understand the nature of the IDs in the deterioration of learning abilities with advancing age. The fact that the procedures used in the present study are differentially sensitive to various sources of IDs even within a group of college-age Ss suggests that these methods have considerable promise for the study of the effects of aging on learning and memory.

3. Backward memory span was found to differ factorially from forward memory span in the first-order analysis. Forward and backward span have much in common, however, and they come together in the higher-order analyses. Unlike forward memory span, backward memory span involves a first-order factor identified as susceptibility to response competition.

4. In the first-order factor analysis serial learning stands quite apart from memory span, although the factors that involve these two types of tests are correlated, so that serial learning and memory span actually share some of the same sources of variance, which are described in terms of the higher-order factors. Because of doublets created by the high correlation between errors and oscillation scores in serial learning, the first-order factors involving serial learning were not very enlightening. The fact that serial learning changes in factorial composition as a result of practice also adds to the difficulty of interpreting the first-order factors. The second-order factors are reasonably clear with respect to serial learning, however. It appears that IDs in serial learning also involve resistance to interference with consolidation of the type found in memory span tasks, and in addition, resistance to interference due to response competition. The Stroop interference score is heavily loaded on this response competition factor which seems to play an important part in the serial learning tasks. This kind of interference seems to operate in situations in which there is competition or mutual blocking among associations or response tendencies that have already been consolidated to some degree. It is a quite different factor from interference with consolidation. Serial lists composed of words and high frequency trigrams were more highly loaded on this factor than lists composed of color-forms, which had somewhat greater loadings on the factor of interference with consolidation. It was hypothesized that extra-experimental sources of interference involving response competition play a greater role in the learning of words and trigrams than in color-forms, and this might be the reason that serial learning of words and trigrams was more heavily loaded on the response competition factor. The memory span tasks, unlike the serial learning tasks, were not loaded on the response competition factor to any appreciable degree.

College grade point average (GPA) had its highest loading on the response competition factor. GPA had practically negligible loadings on all the other factors.

5. Of the two personality factors, Extraversion played a more prominent role in the present experiment than Neuroticism. There is evidently some common genotype underlying extraversion and learning ability, but the nature of this genotype is not clear from the present analysis. The trait

of extraversion, as measured by the Maudsley and Eysenck Personality inventories, was positively correlated with superior performance on most of the tasks in the present battery. Extraversion had a loading of .41 on the general (fourth-order) factor. But the modus operandi of extraversion in the domains of memory span and serial learning remains obscure.

It might be asked whether one is bound to find factors or ID genotypes corresponding to every kind of experimental variation of a learning task. The present study shows that significant sources of IDs are not associated with every kind of experimental variable. For example, it was found that in the normal college population the sensory modality (visual or auditory) of the stimuli is not a significant source of IDs in memory span for either immediate or delayed recall. This is true even though sensory modality has statistically highly significant overall main effects. Auditory digit span is greater than visual for immediate recall, and the reverse is true for delayed recall. But there is no Ss X Modality interaction; that is, the correlation between visual and auditory memory span is perfect. Another instance of the same sort of thing is involved in the pre and post cued conditions of the Forward and Backward memory span tests. Though pre and post cuing for direction of recall of a lengthy digit series make for large and highly significant differences in level of performance, there is not a trace of this experimental variable to be found in the factor analyses. The variable thus seems to be totally unimportant from the standpoint of IDs. On the other hand, the absolute magnitude of retroactive and proactive inhibition differed hardly at all, and yet RI and PI came out on different factors. All this goes to show that genotypic sources of IDs cannot be discovered merely by means of task analysis. It is one thing to analyze a learning task psychologically into what seem to be the elemental processes or components involved in the task and quite another thing to demonstrate that these components are associated with different sources of IDs. The literature on IDs in learning often confuses these two types of analysis or fails to distinguish between them.

IDs in relearning are a puzzle in the present analyses. It seems reasonably clear, however, that relearning involves a different factor or factors than original learning. Even when original learning (OL) has been severely retroactively inhibited so that relearning (RL) takes almost as many trials as OL, there is no significant correlation between RL and OL. This is a very important problem, because in practical situations much of what we call learning actually involves relearning. It would seem most surprising if such relearning involved almost totally different sources of IDs than the original learning. Yet this appears to be the case. Since this result was not anticipated in the present study, not enough additional information was gathered to elucidate the nature of the relearning factor. It was thought that RL after high retroactive interference with OL would involve mainly the factor of response competition, but this seems not to be true. The factorial nature of relearning will have to be elucidated in a specially designed experiment in which adequate account is taken of changes in the factorial composition of serial learning as a function of practice and in which Ss can be equated on degree of OL.

The present study is regarded as a first exploratory step in the investigation of the IDs in the domains of memory span and serial learning. This area of investigation appears to be a fruitful link between the general topics of "intelligence" and "learning ability". It has been shown that memory span and serial learning both involve "learning ability". And what we call learning ability is resolvable into a number of more specific and basic genotypes, whatever their labels might be. If memory span is regarded also as a measure of general intelligence or "capacity", we have shown that this kind of "ability" has factorial structure other than that describable mainly in terms of content, such as verbal, numerical, or spatial ability. The experimental procedures, such as delay in recall, interpolated interference, length of series, intralist similarity, etc., interact with Ss to yield dimensions of IDs that are different from those encountered in traditional psychometric tests. The degree to which psychometric tests overlap the same variance awaits future demonstration; but certainly the kinds of factors we have described in the present study, with the possible exception of the third and fourth order factors, would not be identifiable in a factor analysis of psychometric tests. It seems likely that the kinds of factors measured by the present battery of tests come closer to the kinds of factors that operate in many actual learning situations than do the kinds of knowledge or skills tapped by traditional tests of mental ability, especially where new learning is involved. Achievement-type intelligence tests are predictive of performance in school learning, for example, largely because they indicate how much the individual has been able to learn from his environment in the past and how much information or skill he possesses that can transfer to the learning of related material. IDs in the process of learning itself, however, are not elucidated by achievement-type tests, except perhaps in so far as these assess sources of transfer to the learning of a specific subject matter. Both kinds of tests - those that tap already acquired knowledge and skill as well as those that measure the various factors involved in current basic learning ability - are of course needed for a comprehensive description of an individual's educational potential or for the diagnosis of specific learning disabilities. Tests that tap IDs in the processes of learning itself, however, have not yet been developed. The present study might be regarded as a step in this direction.

Perhaps more important than any immediate diagnostic applications of such tests is the potential contribution that methods similar to those of the present study can make to our theoretical understanding of the nature of the learning process. These theoretical implications of the study of IDs in learning were discussed at length in the first section of this Report This study has brought into focus a number of fundamental questions, hypotheses, and findings concerning IDs in learning. The effort will have been worthwhile to the extent that the methods, findings, and hypotheses offered here are thought worthy of being followed up, to be consolidated or revised and extended for future investigations.

References

Allison, R. B. Learning parameters and human abilities. Unpublished doctoral dissertation. Princeton University, 1960.

Averbach, E., & Sperling, G. Short term storage of information in vision. In C. Cherry (Ed.) Information theory. London: Butterworths, 1961.

Brener, L. R. An experimental investigation of memory span. J. exp. Pshcol., 1940, 26, 467-482.

Broadbent, D.E. Perception and communication. London: Pergamon, 1958.

Broadbent, D. E. Flow of information within the organism. J. verb. Learn. verb. Behav., 1963, 2, 34-39.

Brown, J. Short-term memory. Brit. Med. Bull., 1964, 20, 8-11.

Cameron, D.E. Impairment of the retention phase of remembering. Psychiat. Quart., 1943, 17, 395-404.

Cattell, R. B. Factor analysis. New York, Harper, 1952.

Conrad, R. Very brief delay of immediate recall. Quart. J. exp. Psychol., 1960, 12, 45-47.

Cronbach, L. J. The two disciplines of scientific psychology. Amer. Psychol., 1957, 12, 671-684.

Eysenck, H.J. The dynamics of anxiety and hysteria. London: Routledge & Kegan Paul, 1957.

Eysenck, H.J. Manual of the Maudsley Personality Inventory. London: Univ. London Press, 1959.

Eysenck. H.J. The structure of human personality. London: Methuen, 1960.

Eysenck, H.J. Extraversion and the acquisition of eyeblink and GSR conditioned responses. Psychol. Bull., 1965, 63, 258-270.

Eysenck, H.J. , & Eysenck, S.B.G. Manual of the Eysenck Personality Inventory. London: Univer. London Press, 1964.

Ferguson, G.A. On learning and human ability. Canad. J. Psychol., 1954, 8, 95-112.

Ferguson, G.A. On transfer and the abilities of man. Canad. J. Psychol., 1956, 10, 121-131.

Fleishman, E.A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), Training research and education. Pittsburgh: Univ. of Pittsburgh Press, 1962.

Fleishman, E.A., & Hempel, W.E., Jr. Factorial analysis of complex psychomotor performance and related skills. J. appl. Psychol., 1956, 40, 96-104.

Gagne, R.M. The conditions of learning. New York: Hold, Rinehart, & Winston, 1965.

Games, P.A. A factorial analysis of verbal learning tasks. J. exp. Psychol., 1962, 63, 1-11.

Garrett, H.E. The relation of tests of memory and learning to each other and to general intelligence in a highly selected adult group. J. educ. Psychol., 1928, 19, 601-613.

Gates, A.I. The mnemonic span for visual and auditory digits. J. exp. Psychol., 1916, 1, 393-403.

Gates, A.I., & Taylor, G.A. An experimental study of the nature of improvement resulting from practice in a mental function. J. educ. Psychol., 1925, 16, 583-592.

Gerard, R.W. The material basis of memory. J. verb. Learn. verb. Behav., 1963, 2, 22-33.

Gilbert, L.C. Speed of processing visual stimuli and its relation to reading. J. educ. Psychol., 1959, 55, 8-14.

Gladis, M., & Braun, H.W. Age differences in transfer and retroaction as a function of intertask response similarity. J. exp. Psychol., 1958, 55, 25-30.

Glaser, R. Implications of previous work on learning and individual differences. In R.M. Gagne (Ed.) Learning and individual differences. New York: Bobbs-Merrill, in press.

Guilford, J.P. The difficulty of a test and its factor composition. Psychometrika, 1941, 6, 67-77.

Guilford, J.P., & Dallenbach, K.M. The determination of the memory span by the method of constant stimuli. Amer. J. Psychol., 1925, 36, 621-628.

Hendrickson, A.E., & White, P.O. Promax: A quick method for rotation to oblique simple structure. Brit. J. stat. Psychol., 1964, 17, 65-70.

Hovland, C.I. Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired-associate learning. J. exp. Psychol., 1939, 25, 622-633.

Jensen, A.R. The Maudsley Personality Inventory. Acta Psychol., 1958, 14, 314-325.

Jensen, A.R. Extraversion, neuroticism, and serial learning. Acta Psychol., 1962, 20, 69-77.

Jensen, A.R. Learning abilities in retarded, average, and gifted children. Merrill-Palmer Quart., 1963, 9, 123-140.

Jensen, A.R. Transfer between paired-associate and serial learning. J. verb. Learn. verb. Behav., 1963, 1, 269-280.

Jensen, A.R. Rote learning in retarded adults and normal children. Amer. J. ment. Defic., 1965, 69, 828-834.

Jensen, A.R. Scoring the Stroop Test. Acta Psychol., 1965, 24.

Jensen, A.R., Collins, C.C., & Vreeland, R.W. A multiple S-R apparatus for human learning. Amer. J. Psychol., 1962, 75, 470-476.

Jensen, A.R., & Roden, A. Memory span and the skewness of the serial-position curve. Brit. J. Psychol., 1963, 54, 337-349.

Jensen, A.R., & Rohwer, W.D., Jr. The Stroop Color-Word Test: a review. Acta Psychol., in press.

Langdon, L.N. A two-factor study of simple motor tests with particular reference to practice and prognosis. Brit. J. Psychol., 1937, 27, 411-424.

MacKay, G.W.S., & Vernon, P.E. The measurement of learning ability. Brit. J. educ. Psychol., 1963, 33, 177-186.

Macworth, Jane F. Interference and decay in very short-term memory. J. verb. Learn. verb. Behav., 1964, 3, 300-308.

Madsen, M.C. Distribution of practice and level of intelligence. Psychol. Rep., 1963, 13, 39-42.

Manning, W.H., & DuBois, P.H. Correlational methods in research on human learning. Perceptual and motor skills, 1962, 15, 287-321.

McGeoch, J.A., & Irion, A.L. The psychology of human learning. 2nd ed. New York: Longmans, Green, 1952.

McGhie, A., Chapman, J., & Lawson, J.S. Changes in immediate memory with age. Brit. J. Psychol., 1965, 56, 69-75.

Melton, A.W. Implications of short-term memory for a general theory of memory. J. verb. Learn. verb. Behav., 1963, 2, 1-21.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.

Noble, C.E. Verbal learning and individual differences. In C.N. Cofer (Ed.) Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Postman, L. The present status of interference theory. In C.N. Cofer (Ed.) Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Rapier, Jacqueline L. Measured intelligence and the ability to learn. Acta Psychol., 1962, 20, 1-17.

Raven, J.C. *Guide to the Standard Matrices*. London: H.K. Lewis, 1958.

Spence, K.W. Anxiety (drive) level and performance in eyelid conditioning. *Psychol. Bull.*, 1964, *61*, 129-139.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, *74*, No. 498.

Spies, C.J. A bibliography of factor analytic studies of learning. In P.H. DuBois, W.H. Manning, Y C.J. Spies (Eds.) Factor analysis and related techniques in the study of learning. *Tech. Rep. No. 7,* Contract No. Nonr 816(02), Office of Naval Research and Washington University, 1959.

Stake, R.E. Learning parameters, aptitudes, and achievements. *Psychometric Monogr.*, 1961, No. 9.

Stroop, J.R. Studies of interference in serial verbal reactions. *J. exp. Psychol.*, 1935, *18*, 643-661.

Stroud, J.B., & Carter, L.J. Inhibition phenomena in fast and slow learners. *J. educ. Psychol.*, 1961, *52*, 30-34.

Stroud, J.B., & Schoer, L. Individual differences in memory. *J. educ. Psychol.*, 1959, *50*, 285-292.

Thurstone, L.L., & Thurstone, Thelma G. Factorial studies of intelligence. *Psychometric Monogr.* No. 2, 1941.

Underwood, B.J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, 45, 143-149.

Underwood, B.J. Speed of learning and amount retained. *Psychol. Bull.*, 1954, *51*, 276-282.

Underwood, B.J., & Schulz, R.W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

Wechsler, D. *Manual of the Wechsler Intelligence Scale for Children*. New York: Psychological Corp., 1949.

Wechsler, D. *The measurement and appraisal of adult intelligence*. Fourth edition. Baltimore: Williams & Wilkins, 1958.

Welford, A.T. *Aging and human skill*. London: Oxford Univ. Press, 1958.

Wickelgren, W.A. Acoustic similarity and retroactive interference in short-term memory. *J. verb. Learn. verb. Behav.*, 1965, *4*, 53-61.

Woodrow, H. The effect of practice on groups of different initial ability. *J. educ. Psychol.*, 1938, *29*, 268-278.

Woodrow, H. The ability to learn. Psychol. Rev., 1946, 53, 147-158.

Wooldridge, D.E. The machinery of the brain. New York: McGraw-Hill, 1963.

Young, J.Z. A model of the brain. London: Oxford Univ. Press., 1964.

Appendix A

Analysis of Variance of Visual and Auditory Memory Span Tests (IDS and DDS, 2-9) Based on Position Scores (N = 50)

Source of Variance	df	Mean Squares	
		Visual	Auditory
1. Conditions	1	304.98	2075.66
2. Days	1	268.64	64.62
3. Replications	4	73.03	15.46
4. Series Lengths	7	1534.81	1188.55
5. Subjects	49	48.72	42.97
12	1	19.41	12.11
13	4	15.13	10.56
14	7	25.22	185.49
15	49	6.61	8.71
23	4	6.67	3.51
24	7	24.89	14.19
25	49	6.70	3.62
34	28	11.06	7.83
35	196	2.02	2.20
45	343	7.63	6.61
123	4	6.94	1.86
124	7	1.76	1.21
125	49	1.19	2.08
134	28	3.62	4.05
135	196	1.62	1.55
145	343	1.76	3.10
234	28	3.98	6.33
235	196	1.84	1.68
245	343	2.01	2.09
345	1372	1.50	1.67
1234	28	4.57	4.49
1235	196	1.57	1.20
1245	343	1.44	1.65
1345	1372	1.48	1.41
2345	1372	1.40	1.55
Residual	1372	1.42	1.71
Total	7999		

Appendix B

Analysis of Variance of Color-Form Memory Span Tests (ICFS and DCFS, 2-9), Based on Position Scores (N = 30)

Source of Variance	df	Mean Squares
1. Conditions	1	361.90
2. Days	1	226.20
3. Replications	4	25.75
4. Series Lengths	7	24.55
5. Subjects	29	33.09
12	1	0.75
13	4	5.29
14	7	13.01
15	29	2.48
23	4	5.36
24	7	3.31
25	29	3.41
34	28	4.94
35	116	1.42
45	203	1.81
123	4	1.70
124	7	2.96
125	29	0.98
134	28	5.00
135	116	1.23
145	203	1.06
234	28	1.26
235	116	1.17
245	203	1.16
345	812	1.00
1234	28	0.99
1235	116	0.81
1245	203	1.02
1345	812	1.03
2345	812	0.99
Residual	812	0.94
Total	4799	

Appendix C

Analysis of Variance of Forward-Backward Memory Span Tests, Based on Percentage of Maximum Possible Score, for Position and Sequence Scores (_N_ = 30)

Source of Variance	df	Mean Squares Position Scores	Mean Squares Sequence Scores
1. Cuing (Pre vs. Post)	1	232221.19	65438.46
2. Days	1	15657.31	46550.82
3. Replications	15	1065.74	1250.66
4. Direction (F vs. B)	1	41002.74	11703.95
5. Subjects	29	13816.00	7018.67
12	1	9.83	320.14
13	15	300.30	332.13
14	1	1647.32	1614.88
15	29	10917.88	778.74
23	15	488.87	490.82
24	1	2272.77	1921.18
25	29	818.18	613.98
34	15	373.29	186.90
35	435	258.28	288.22
45	29	984.71	4554.26
123	15	158.06	242.07
124	1	1236.81	1573.25
125	29	1174.52	417.55
134	15	505.26	485.35
135	435	262.02	319.13
145	29	1231.89	926.14
234	15	503.40	102.57
235	435	239.79	320.15
245	29	337.42	1137.60
345	435	237.74	290.87
1234	15	294.06	128.44
1235	435	206.47	264.93
1245	29	407.55	351.76
1345	435	238.25	277.03
2345	435	207.12	247.58
Residual	435	233.51	289.92
Total	3839		

Appendix D

Analysis of Variance of Digit Span and Repeated Digit Span Tests (DS and RDS, 5-12), Based on Position Scores. ($\underline{N}$ = 30)

Source of variance	df	Mean Squares
1. Conditions (DS vs. RDS)	1	340.79
2. Days	1	481.96
3. Replications	4	119.08
4. Length of Series	7	118.97
5. Subjects	29	159.40
12	1	5.96
13	4	5.17
14	7	8.80
15	29	7.84
23	4	47.44
24	7	39.30
25	29	11.99
34	28	6.69
35	116	5.07
45	203	10.44
123	4	2.48
124	7	1.51
125	29	4.26
134	28	4.93
135	116	2.21
145	203	2.69
234	28	4.57
235	116	4.31
245	203	3.27
345	812	2.61
1234	28	2.50
1235	116	2.34
1245	203	2.80
1345	812	2.56
2345	812	2.57
Residual	812	2.66
Total	4799	

Appendix E

Factor Loadings Corrected for Attenuation (Decimals omitted)

Reliability	20	34	40	50	60	70	80	90	95
20	45	67	89	100					
30	36	55	73	91	100				
40	32	47	63	79	95	100			
50	28	42	56	71	85	99	100		
60	26	39	52	64	77	90	100		
70	24	36	48	60	72	84	96	100	
80	22	34	45	56	67	78	89	100	
90	21	32	42	53	63	74	84	95	100
95	21	31	41	51	62	72	82	92	97

Appendix F

Estimated Standard Error of Factor Loadings After Correction for Attenuation (Decimals Omitted)

	Factor Loading (Corrected for Attenuation)								
Reliability	20	30	40	50	60	70	80	90	95
50	02	03	04	05	06	07	08	10	10
60	02	02	03	04	04	05	06	07	97
70	01	02	02	02	03	04	04	05	05
80	01	01	0	02	02	02	02	03	03
90	00	00	01	01	01	01	01	01	01
95	00	00	00	00	00	01	01	01	01

Correlations of the Maudsley Personality Inventory With Memory Span Tests in the Single-Test Groups (Decimals omitted)

				Correlations	
Test		Day	N[1]	Neuroticism	Extraversion
1.	IDS, 2-9	1	50	01	05
2.	"	2	"	-15	07
3.	DDS 2-9	1	"	-03	21
4.	"	2	"	-12	23
5.	ICFS, 2-9	1	30	10	46
6.	"	2	"	-10	28
7.	DCFS, 2-9	1	"	-02	42
8.	"	2	"	00	36
9.	DS-RI, 4-7	1	52	-24	02
10.	"	2	"	-12	-06
11.	DS-PI, 4-7	1	"	-22	-04
12.	"	2	"	-03	07
13.	DDS, 4-7	1	"	-39	-06
14.	"	2	"	-23	00
15.	IDS, 4-7	1	"	-25	-15
16.	"	2	"	-12	-09
17.	FDS, Pre	1	30	-11	09
18.	"	2	"	-08	07
19.	FDS-Post	1	"	08	03
20.	"	2	"	-04	05
21.	BDS-Pre	1	"	17	-14
22.	"	2	"	00	-16
23.	BDS-Post	1	"	-04	11
24.	"	2	"	02	16
25.	DS, 5-12	1	"	12	04
26.	"	2	"	08	-04
27.	RDS, 5-12	1	"	12	-06
28.	"	2	"	11	01

---

[1] Dittos indicate that the particular test was taken by the same group of Ss that took the preceding test.